

请参阅本出版物的讨论、统计数据和作者简介: <https://www.researchgate.net/publication/330893507>


云中负载均衡技术的问题和挑战 计算:一项调查

ACM计算调查中的文章 · 2019 年 2 月
DOI:10.1145/3281010

引文
185

阅读
10,767

2位作者,包括:



帕万·库马尔

印度昌迪加尔国家技术教师培训和研究院

6 篇出版物 383次引用

查看资料

云计算中负载均衡技术的问题和挑战:调查

PAWAN KUMAR 和 RAKESH KUMAR, NITTTR,昌迪加尔

随着计算技术的发展,云计算为用户服务增加了一种新的范式,允许在任何时间、任何地点以按次付费的方式访问信息技术服务。

由于云服务的灵活性,许多组织正在将业务转移到云端,服务提供商正在建立更多的数据中心来为用户提供服务。然而,提供具有成本效益的任务执行和资源的正确利用是至关重要的。文献中报道了几种基于负载均衡、任务调度、资源管理、服务质量和 workload 管理来提高性能和资源使用的技术。云中的负载均衡可以让数据中心避免虚拟机过载/欠载,这本身就是云计算领域的一个挑战。因此,开发人员和研究人员有必要为并行和分布式云环境设计和实现合适的负载均衡器。这项调查对与现有负载均衡技术相关的问题和挑战进行了最先进的回顾,以便研究人员开发更有效的算法。

120

CCS 概念: 计算方法→分布式算法;并发算法;

其他关键词和短语:负载均衡、云计算、资源分配、任务调度、虚拟机、工作负载管理、优化

ACM参考格式:

帕万·库马尔和拉克什·库马尔。2019。云计算中负载均衡技术的问题和挑战:调查。ACM 计算。幸存者。51:6,第120条(2019年2月),35页。<https://doi.org/10.1145/3281010>

1 简介

随着信息技术(IT)的快速发展,云计算已经成为传统计算技术的替代品,可以随时随地为客户提供按使用付费的服务(Brown 2017;Buyya et al. 2010),它使用户能够访问可配置的计算资源池(服务器、存储、网络、应用程序)。多家公司(例如 Amazon Web Services (AWS)、Microsoft Azure、Google、IBM 云、Rackspace、Red Hat、Verizon 云、VMware) (Technavio 2018) (也称为云计算提供商)向用户提供这些云服务。云的主要目的是利用分布式资源

作者地址:P. Kumar 和 R. Kumar,计算机科学与工程系,国家技术教师培训与研究学院,Sector-26,昌迪加尔,印度;电子邮件:pawan.cse@nitttrchd.ac.in,raakeshdhiman@gmail.com。

允许免费制作本作品全部或部分内容的数字或硬拷贝以供个人或课堂使用,前提是制作或分发副本不是为了盈利或商业利益,并且副本在首页上附有此通知和完整引用。必须尊重 ACM 以外的其他人拥有的本作品组件的版权。

允许以信用方式提取。要以其他方式复制、重新发布、发布到服务器上或重新分发到列表,需要事先获得特定许可和/或付费。从 permissions@acm.org 请求权限。© 2019 计算机协会。

0360-0300/2019/02-ART120 15.00 美元
<https://doi.org/10.1145/3281010>

有效地获得高吞吐量和性能。它使云能够解决需要高计算能力的问题。它还允许分配所有资源在世界各地的不同数据中心执行任务,为客户提供更快的服务

(达斯古普塔等人, 2013)。连同云计算分发系统 (Milani 和 Navimipour 2016年; Navimipour 2015), 还有许多其他分布式系统可用,例如网格计算

(Khanli 等人, 2008 年)和点对点计算 (Navimipour 和 Milani, 2015 年), 它们允许资源共享和数据传输设施。这为云服务提供商建立新的数据中心和云用户将其业务置于经济高效的云上提供了商机。云计算允许用户动态扩展或扩展虚拟资源

根据计算要求无需人工交互 (Chiregi 和 Navimipour 2016)。

云计算可以按两种方式分类:基于位置或提供的服务。基于在位置上,云可以是公共云、私有云、混合云或社区云。公共云服务可供任何人使用,并且基础设施位于服务提供公司的场所内。

公有云最容易受到各种攻击,但成本效益最高。私人

云专供特定用户或组织使用。它提供最高的安全性

并且控制水平给用户带来了更高的成本。混合云是公共云和私有云的组合,根据组织需求用于不同的目的。一个社区

云由许多组织使用的通用基础设施组成,这些组织共享数据和

管理。基于服务,云主要分为基础设施即服务 (IaaS)、

平台即服务 (PaaS) 或软件即服务 (SaaS)。在 IaaS 中,云提供基础 IT

资源,例如网络功能、计算机、更大的灵活性以及对计算资源的控制。PaaS 消除了组织处理基本基础设施的需要 - 通常,

计算机硬件和操作系统 (OS), 让您可以专注于部署

应用程序。SaaS 允许用户专注于特定软件的使用而不是思考

关于如何管理基础设施和服务。除了这些服务之外,云计算还提供数据库即服务 (DaaS) (Hacigumus 等人, 2002 年)、专家即服务 (EaaS)

(Jafari 等人, 2015)、存储即服务 (SaaS)、网络即服务 (NaaS)、安全即服务

(SECaaS) (Candrlic 2013)、通信即服务 (CaaS)、监控即服务 (MaaS)、

和测试即服务 (TaaS) (Dhillon 2015), 用户可以出于不同的应用程序目的访问它们。许多云应用程序可用于教育的不同方面,

健康监测 (Punj 和 Kumar 2018)、数据分析和机器人技术。用户无需了解其计算环境所需的技术即可访问它们。

根据班纳吉的说法

等人。(2015), 云还提供了灵活性和可扩展性来发布和获取各种

可根据应用程序需求配置资源。

由于云提供不同的服务,因此需要服务质量 (QoS) 监控来评估

为满足用户需求并维护服务级别协议 (SLA) 而提供的服务。

在此过程中,云可能会带来一些问题和挑战,例如负载均衡

(Chen 等人, 2017 年; Mohamed 等人, 2013 年; Nuaimi 等人, 2012 年; Pacini 等人, 2015 年; Suresh 等人, 2014 年), 性能分析和建模 (Garg et al. 2013; Ghosh et al. 2013; Li et al. 2012; Mauch et al. 2012)

2013 年; 米格尔等人。2015 年; 苏萨等人。2015)、吞吐量和响应时间 (Pacini 等人, 2015)、安全和隐私问题 (Khan 等人, 2013; Khorshed 等人, 2012; Lombardi 和 Di Pietro, 2011;

苏巴西尼和卡维莎 2011; Zissis 和 Lekkas 2012)、资源管理 (Jennings 和 Stadler

2015 年; 马里内斯库等人。2017) 和服务质量。目前,云负载均衡 (LBC) 是其中之一

避免任务计算期间虚拟机过载/欠载情况的主要挑战。因此,有必要确定影响 LBC 和

为云环境开发有效的负载均衡技术。

本研究的主要目标如下:

- 研究文献中现有的各种负载均衡技术。 · 对各种负载均衡技术进行分类并概述现有的挑战
- 和
- 负载均衡的问题。 · 概述未来
- 改进负载均衡技术的研究领域。

1.1 负载均衡的需求负载均衡提供了在可

用资源上平均分配工作负载的工具。其目标是通过配置和取消配置应用程序实例以及正确利用资源,在任何服务组件发生故障时提供连续服务。此外,负载均衡的目的是最大限度地缩短任务的响应时间,提高资源利用率,从而以较低的成本增强系统性能。负载均衡还旨在为那些未来规模可能会增加并需要更多资源的应用程序提供可扩展性和灵活性,并为需要即时执行的作业提供与其他作业相比的优先级。负载均衡的其他目标包括减少能源消耗和碳排放、避免瓶颈、资源配置以及满足改善负载均衡的 QoS 要求。需要考虑不同指标的适当的工作负载映射和负载均衡技术。

1.2 动机

- 云中的负载均衡是在虚拟机上平均分配工作负载以正确利用资源的过程。在本次调查中,根据不同的指标强调了各种负载均衡算法。
- 负载均衡器有助于以最小的成本公平地将资源分配给任务,从而实现资源利用率和用户满意度,这激励我们发现负载均衡中的问题并致力于解决它们。 · 基于云计算领域的持续需求和不断增加的工作量,我们认识到云资源之间负载均衡的需求。因此,在现有研究的基础上,系统
- 地识别和总结了现有的工作,描绘了未来研究工作的问题和挑战。

1.3 文章组织本文的组织方式如下:第

2节讨论之前针对云中的负载均衡所做的调查。第3节描述了查找可用研究论文所采用的调查技术、搜索标准、信息来源和质量评估。第4节介绍了负载均衡方面的各种挑战。第5节讨论负载均衡的分类、度量 and 策略。第6节包含各种负载均衡技术的文献综述以及用于演示这些负载均衡和资源调度工具的一些模拟工具。在第7节中,我们对所提到的各种技术进行了讨论。第8节描述了现有负载均衡技术面临的问题和挑战。我们在第9节中提出结论并阐述未来趋势。

2 相关调查

全球研究界对设计和开发最佳资源利用技术表现出了兴趣,因为当前的研究和评论正在引起他们的注意。云中的负载均衡是在未提交的虚拟机上分配执行负载以提高系统吞吐量的过程。响应时间和执行成本最小化 (Goyal

和 Verma 2016) 需要提高云计算环境的成功率。随着负载均衡方面存在各种挑战,例如云中的资源调度、性能监控、QoS 管理、能耗和服务可用性 (Kaur 和 Luthra)

2012年;马拉迪2015)。在云计算、负载均衡、安全、能耗、资源管理等方面已经做出了大量的研究工作。负载均衡和任务

调度是主要问题,需要更多研究。我们提供全面的审查

云的各类调度、负载均衡、基于任务的负载均衡技术

在本文中。在本节中,我们回顾了一些对负载均衡技术进行了深入调查的文章。

Ghomi 等人对云中的负载均衡进行了广泛的审查。(2017),谁

将评论分为七个不同的类别:Hadoop MapReduce 负载均衡技术、

基于自然现象的负载均衡技术、基于Agent的负载均衡技术、

一般负载均衡技术、面向应用的负载均衡技术、网络感知任务调度和负载均衡、工作流特定调度算法他们讨论了现有负载均衡技术中的各种问题和挑战以及未来的方向。他们主要关注 Hadoop MapReduce 和能源效率,这

关注云负载均衡;尽管如此,他们的工作缺乏基于任务和基于集群的负载均衡,这也是在可用资源上执行任务的重要问题。因此,有

需要确定当今高度集成的基于任务和基于集群的负载均衡中的问题

可扩展且高度分布式的计算环境。

Milani 和 Jafari (2016)回顾了各种现有的负载均衡方案并对它们进行了分类

分为动态和混合子域。他们描述了这些技术的行为基于

不同的参数,以及它们的优点、缺点和挑战。他们表示

解决这些算法的问题,以开发更有效的算法,以最大限度地减少资源消耗和功耗,并使负载均衡技术更加有效。然而,他们没有讨论基于任务的负载均衡、基于集群的负载均衡和能源

消费问题。

辛格等人。(2017)对基于元启发式的任务调度算法进行了广泛的综述,而 Kalra 和 Singh (2015) 的综述仅限于元启发式技术

仅用于工作流程调度。作者辛格等人。(2017)确定了相关的各种问题

云中的任务调度,并基于元启发式方法对依赖任务和独立任务进行了比较分析。在辛格等人中。(2017)以及卡尔拉和辛格

(2015)大多数文献集中于遗传算法、粒子群优化

(PSO)、基于蚁群优化(ACO)的调度技术,缺乏基于任务的负载均衡的覆盖,这也是目前云中的一个主要问题。辛格等人。(2017)还

尚未评估其他主要问题,例如碳排放、基于集群的负载均衡和

最小的资源消耗。

Singh 和 Chana (2015)对自动云资源管理进行了审查,

他基于六种不同的观点(资源分配、工作负载调度、监控、QoS 要求、应用程序设计、自我管理)提出了这项研究,并讨论了基于不同特征的分类法。他们专注于改善管理和

资源的使用,而任务执行时间和响应时间对于提供 QoS 更为重要同时履行 SLA。

基于指标(例如

分布式环境、集群中的响应时间、迁移时间、可扩展性和资源使用)

Ivanisenko 和 Radivilova 提出了分布式云计算环境中的系统

(2015)。作者介绍了最常用的特征、优点和缺点

负载均衡算法,但忽略了现有技术中的问题、挑战和未来趋势。

Katyal 和 Mishra (2014) 根据 SLA 中指定的用户需求对各种云环境的负载均衡算法进行了分析。作者根据不同的类别回顾了这些技术,包括任务依赖性、节点的地理分布和云环境。他们讨论了主要类别中现有技术的优点、缺点和挑战,但缺乏基于各种负载均衡参数的评估。

根据分析,我们观察到,没有对任务调度和负载均衡技术进行全面的审查,以展示这些技术的重要性、分类和未来问题。在本文中,我们制定了一系列问题来选择最重要的研究论文进行审查,并对它们进行分类以回答这些问题。

3 调查技术在本文中,我们遵循了

Kitchenham (2004)、Kitchenham 等人提出的不同领域的各种指南。(2009),库皮艾宁等人。(2015)、Charband 和 Navimipour (2016)以及 Navimipour 和 Charband (2016)进行了系统调查并重点关注云负载均衡相关研究。我们制定了审查方法,研究了方法,得出了结果,探索了挑战。在本节中,我们讨论研究文章的信息来源、选择标准、质量评估和结果评估。表1包含我们的研究问题列表,用于规划负载均衡调查并确定当前与负载均衡相关的问题。图1显示了用于根据包含-排除标准搜索和识别文章的七个阶段过程。

3.1 信息来源

我们在 Scopus、Web of Science、Google Scholar、书籍和杂志中广泛搜索期刊和会议研究文章,作为提取相关文章的数据来源。我们的搜索中使用了以下数据库:

- IEEE Xplore (<http://ieeexplore.ieee.org>) ·
- 施普林格(<https://link.springer.com>) ·
- ScienceDirect (<http://www.sciencedirect.com>) ·
- 谷歌学术(<https://scholar.google.co.in>) · Scopus
- (<https://www.scopus.com>) · ACM 数
- 字图书馆(<https://www.acm.org/digital-library>) · 泰勒和弗朗西斯
- (<https://www.taylorandfrancis.com>)

在图2 中,我们显示了来自不同来源的审阅论文的百分比。

3.2 检索标准

我们定义了在上述数据库中搜索的关键词。每次搜索的摘要中都涉及关键词“负载均衡”和“云”。这是一种常见的方法,而且耗时。我们搜索了与结果相匹配的不同同义词和相关关键词,包括“任务迁移”、“资源利用”、“分布式”、“虚拟机迁移”、“工作负载”和“集群”。我们再次细化查询以匹配具体结果,并再次应用于研究论文的标题、摘要。我们于 2010 年 1 月至 2018 年 1 月期间进行了这项研究。

表 1. 复习问题

不。	研究问题	动机
1.	负载均衡在网络中的意义是什么云？	主要目的是识别各种负载均衡研究/发表的文章随着使用的增加,时间及其重要性云端。多种负载均衡技术目前已提出,需要基于各种负载均衡进行评估指标。这项调查还旨在确定现有负载均衡中的问题和挑战确保基于 QoS 的服务的技术。这里,我们讨论各种负载均衡技术根据不同的类别。
2.	云中为什么需要负载均衡？	
3.	现有技术能否实现负载均衡基本指标？	
4.	目前云负载状况如何平衡？	
5.	选择合适服务的标准是什么为消费者提供服务？	
6.	云之间谈判的标准是什么服务提供商和云消费者？	
7.	如何最大限度地减少电力消耗云数据中心及其影响环境？	
8.	需要哪些新算法建议提高系统性能？	
9.	哪些资源和参数最多对云性能很重要吗？	
10.	一个人如何管理资源来克服困难资源超载问题和负载不足？	
11.	如何减少响应时间并改进资源利用？	
12.	用户对QoS的要求是什么期望使用云？	
13.	如何开发负载均衡技术满足用户的QoS要求?不同的负载均衡类别以揭示重要的	各种研究论文需要从研究问题。随着需求的不断增加随着时间的推移,云计算变得至关重要负载均衡标准并开发技术处理大量用户请求有效地。这里讨论的各种问题将帮助确定未来的研究领域。
14.	如何验证现有的和通过开发负载均衡技术可用的工具？	
15.	如何开发一种架构来满足用户的基本QoS需求是什么？	
16.	SLA 是如何定义的?评判标准是什么定义 SLA 违规规则？	
17.	有哪些仿真工具可用于负载均衡技术的验证？	识别各种云模拟很有用工具将帮助算法开发人员验证负载均衡和资源调度技术。
18.	哪些算法验证参数是这些模拟器考虑过吗？	

从搜索到的论文中,我们将符合 Kitchenham 等人提出的质量评估检查表 (QAC) 的论文纳入我们的调查中。[\(2009\)](#)。它包括来自同行评审的论文期刊、书籍、会议、杂志、研讨会、白皮书和网站。

3.3 质量评估

对于搜索到的文章,我们在纳入和排除后应用了质量评估流程标准如表2所示。我们根据不同的基本关键词发现了1,024篇文章来自各种来源的期刊,包括 IEEE Xplore, Springer, Science Direct 和 ACM。我们排除了一些与我们的研究不相关的标题的文章。读完摘要后，

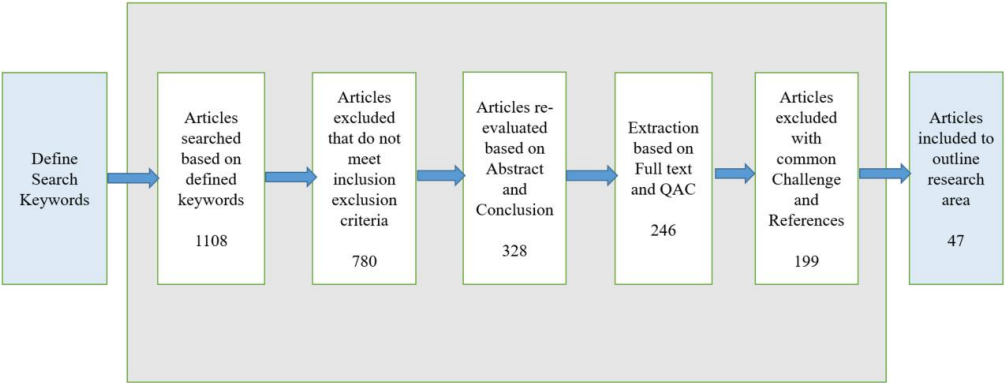


图 1. 物品识别流程概述。

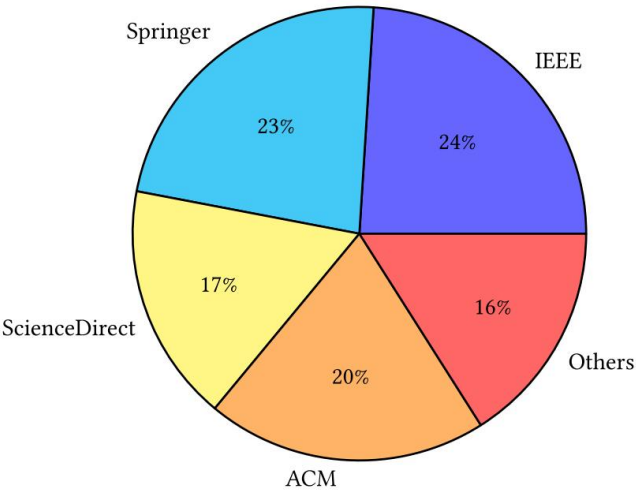


图 2. 调查的负载均衡研究论文来源。

我们排除了一些不符合我们标准的文章。对于其余的论文,我们回顾了完整地查阅了文章,找到了 47 篇研究论文,这些论文纳入了我们的评论。

云中负载均衡的 4 个挑战

云计算技术正在成为数据领域更高级研究的目标以及理论和实践方面的计算。然而,云计算研究面临着很多问题,负载均衡是需要特殊解决的突出挑战之一注意力。此外,还有其他几个问题,例如虚拟机 (VM) 迁移、VM 安全、用户 QoS 满意度和资源使用需要同等关注,以便找到改善云资源使用的最佳解决方案。下面讨论了一些负载均衡问题:

- (1)地理分布式节点:一般来说,云中的数据中心出于计算目的是地理分布的。在这些中心,空间分布的节点是
被视为有效执行用户请求的单一位置系统。一些负载均衡技术是为较小的区域设计的,在这些区域中它们不考虑
网络延迟、通信延迟、分布式之间的距离等因素

表 2. 论文纳入-排除标准

标准1. 从	
包容性	服务提供的角度清楚地描述负载均衡技术。 2. 云计算领域发表。 3. 经过同行评审并用英语撰写。 4. 在知名期刊、会议和杂志上发表文章。 5. 由学术或工业研究人员撰写。
排除	1.不关注云中的负载均衡。 2. 有共同的挑战和参考。

计算节点、用户与资源的距离等。位于非常远的位置的节点是一个挑战,因为这些算法不适合这种环境。因此,应该考虑为远程节点设计负载均衡算法。

(2)单点故障:设计了各种动态负载均衡算法,其中一些技术是非分布式的,负载均衡的决策由中心节点做出。如果中央设备崩溃,那么就会影响整个计算环境。因此,需要开发一些单个节点不控制整个计算系统的分布式算法。

(3)虚拟机迁移:虚拟化允许在一台物理机上创建多个虚拟机。这些虚拟机本质上是独立的并且具有不同的配置。

如果物理机过载,一些虚拟机需要使用虚拟机迁移负载均衡方法转移到远程位置。

(4)异构节点:在云负载均衡的早期研究中,研究人员对同构节点进行了理论分析。在云计算中,用户需求动态变化,需要在异构节点上执行它们,以有效利用资源并最大限度地缩短响应时间。因此,针对异构环境发明高效的负载均衡技术对研究人员来说是一个挑战。

(5)存储管理:云存储解决了传统存储系统较旧的需要人员管理和硬件成本较高的问题。云允许用户异构地存储数据,而不会出现任何访问问题 (Wu et al. 2012) 。

云存储日益增长,这需要存储数据的复制以实现数据的高效访问和一致性。由于复制点上的重复数据存储策略,全量数据复制方案效率不高。部分复制就足够了,但可能存在数据集可用性问题,并且增加了负载均衡技术的复杂性。因此,需要开发一种有效的负载均衡技术,考虑基于部分复制系统的应用程序和相关数据的分布。

(6)负载均衡器的可扩展性:云服务的按需可用性和可扩展性允许用户随时访问服务以快速缩小或扩展。一个好的负载均衡器应该考虑计算能力、存储、系统拓扑等方面需求的快速变化,以有效地促进这些变化 (Ray 和 Sarkar 2012) 。

(7)算法复杂性:在云计算中,算法应该简单且易于实现。复杂的算法会降低云系统的性能和效率。

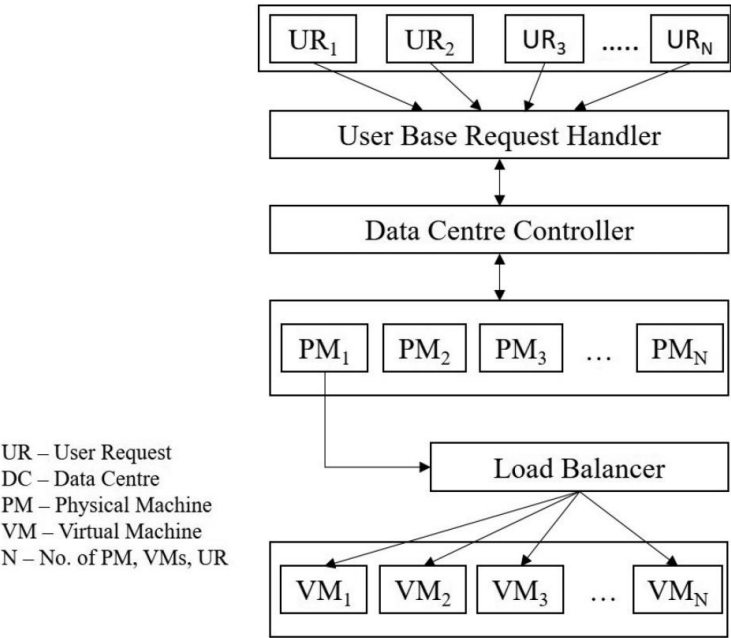


图 3. 负载均衡模型。

5 负载均衡模型、分类、指标和策略云提供对共享资源池（例如服务器、存储和网络（Zeng 等人，2015；Zhang 等人，2010）的按需访问，这需要对用户的工作量和资源进行大量的控制和管理。为了管理用户对可用资源的请求，需要一个良好的负载均衡器来根据虚拟机的 QoS 要求将任务分配给虚拟机（Bhardwaj 和 RamaKrishna 2018）。负载均衡器的模型和工作流程如图 3 所示。云观察到需要动态环境来执行任务的用户请求的很大变化。当云检测到来自用户群的任何请求时，云上的服务代理会识别资源的可用性，并与其他代理协商资源的性能和成本。分析可用资源后，代理将用户请求传输到选定的数据中心，数据中心控制器 (DCN) 接受这些请求以进行进一步处理。数据中心由物理资源/机器组成，它们接收请求并将其传输到服务器上可用的负载均衡器，负载均衡器将任务分配给虚拟机 (VM) 来执行。在此过程中，负载均衡器从状态表中返回虚拟机的可用性，并在分配后更新状态表。如果没有检测到空闲虚拟机，DCN 会将任务放入队列中并等待资源可用。一旦虚拟机完成一个任务，负载均衡器就会将其分配给另一个任务来处理。数据中心还有一个虚拟机管理器，负责处理物理服务器上的所有虚拟机。负载均衡器负责为任务分配合适的虚拟机，其中任务分配是云中的关键问题。负载均衡器还确保虚拟机不会过载或负载不足。如果某些虚拟机负载不足或空闲，而另一些虚拟机过载，则系统性能和服务质量将下降，这可能导致用户放弃服务。虚拟化技术通过在虚拟机之间共享物理资源来帮助使用物理资源。虚拟机管理程序提供虚拟化设施来创建和管理云中的虚拟机。虚拟机管理程序还提供四个重要操作：配置、多路复用、暂停和实时迁移，以实现任务和 VM 的负载均衡（Hwang 等人，2013 年）。A

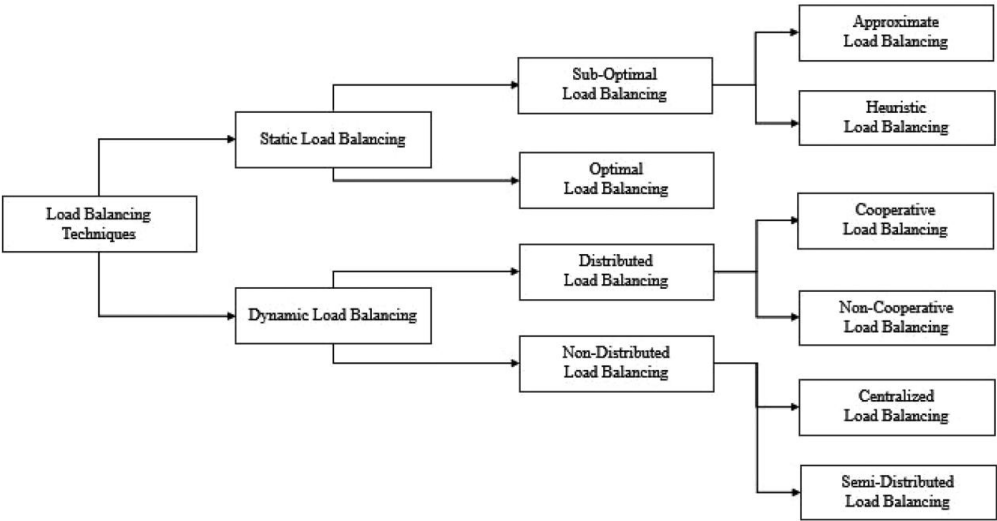


图4. 基于系统状态的负载均衡策略分类。

负载均衡器主要考虑对任务进行调度并将其分配给完美的资源。良好的负载均衡器将提高资源利用率和可用性,并最大限度地缩短任务的响应时间。

5.1 基于系统状态的负载均衡策略的分类负载均衡技术主要根据系统状态和进程启动来分类。基于系统状态的负载均衡方法进一步分为静态和动态两类,如图4所示。根据进程启动,它们分为发送者发起的、接收者发起的和对称的,如下详述。

发送者发起:在这种技术中,如果节点过载,它会寻找其他负载较轻的节点来分担工作负载。当节点拥塞时,发送方启动寻找负载不足的节点的过程。

接收器发起:在这种技术中,接收器或轻负载节点寻找重负载负载节点来分担工作负载。

对称:在该技术中,发送方发起的进程和接收方发起的进程技术相结合。

根据系统状态,负载均衡技术可以分为以下子类别:

静态:静态负载平衡技术遵循一组不依赖于系统当前状态的固定规则。静态算法不灵活,需要先了解资源,例如通信时间、节点的内存和存储容量、节点的处理能力等。这种技术简单易行,但通常无法检测到连接的服务器,从而导致资源分配不均匀 (Chen et al. 2017) 。

该技术的主要问题是决策过程中没有考虑系统的当前状态。因此,它不适合动态改变状态的分布式系统。仅当节点中发生较低的负载波动时,静态技术才能很好地发挥作用。静态负载均衡技术分类如下:

·最优:在最优技术中,DCN 收集有关资源的信息并将任务发送到负载均衡器,负载均衡器在最短时间内进行最优分配。

·次优:如果负载均衡器无法确定最佳决策,它将计算次优解决方案。一些静态技术包括最小-最小-最大-最小-循环-最短作业优先-两阶段机会负载均衡 (OLB) 和虚拟机中央负载均衡。

·动态:这些技术考虑系统的当前状态并在此基础上做出决策。这些技术的主要优点是它们允许将任务从过载的机器转移到负载不足的机器。动态负载均衡技术非常灵活,可以提高系统性能。在处理过程中,动态技术采取以下步骤。它持续监控节点的负载。在给定的时间间隔内,它在节点之间交换负载和状态信息,以计算节点的工作负载并在节点之间重新分配工作负载。如果节点过载,它将负载转移到负载不足的节点。一些负载均衡技术包括基于代理的负载均衡 (Singh 等人, 2015 年)、蜜蜂行为启发负载均衡 (Babu 和 Krishna, 2013 年)、蚁群优化 (Nishant 等人, 2012 年) 和 Throttled (Domanal 和雷迪2013)。进一步的动态负载均衡技术可以分类如下:

·分布式:在分布式技术中,所有节点都参与负载分配,例如任务调度或资源分配 (Cosenza 等人, 2011 年; Shi 等人, 2011 年)。所有节点都维护一个用于通信的信息库,以有效地分配和重新分配任务。分布式算法可以以合作或非合作的形式相互交互。如果系统中的所有节点共同努力实现共同的目标或决策,则称为协作;否则,不合作。

·非分布式:在非分布式技术中,单个节点或一些节点做出负载分配的决定 (Ahmad 和 Ghafoor 1991; Das 等人2003)。非分布式技术在行为上可以是集中式的或半分布式的。在集中式技术中,单个节点执行所有负载分配活动并负责负载均衡。

容错是中心化技术的一个问题,在单节点故障的情况下,节点信息可能会丢失并且可能无法恢复。在半分布式技术中,集群在节点中形成,每个集群作为集中式技术工作。

5.2 负载均衡指标为了更好地利用资源并提高性能,负载均衡器必须将计算负载分散到可用资源上。不同的研究人员提出了几种负载均衡技术和应用于这些技术的各种度量,以提高用户满意度和资源利用率。为了提高整体系统性能,研究人员需要确保所有参数都得到最佳满足。Chen 等人已经讨论了这些指标。(2017)、Daraghmi 和 Yuan (2015)、Abdulhamid 等人。(2014)、Kansal 和 Chana (2012)、Randles 等人。(2010)、Babu 和 Krishna (2013)、Voorsluys 等人。(2011)和拉梅扎尼等人。(2014),如下:

·性能:在实施该技术后,必须与其他现有的负载均衡技术进行比较,验证系统的有效性。

·响应时间:在系统上完成提交的请求所花费的总时间。

·吞吐量:系统上单位时间内完成的提交任务或进程的总量。吞吐量越高,系统性能越好。

·可扩展性:系统在负载均衡时实现统一负载均衡的能力。

所需的节点数量增加。

·容错性:负载均衡技术在任何链路或节点发生故障的情况下仍能一致执行的能力。 ·迁移时间:迁移时间用于计算将请求/任务从过载计算机传输到负载不足的计算机所需的总时间。迁移时间越短,云系统的性能越好。

·资源使用:进行评估以确保所有资源在云系统中得到正确使用。更高的资源使用量将最大限度地降低总体成本,并降低云系统中的能源支出和碳释放率。 ·不平衡程度:这描述了VM之间的差异。 · Makespan:用于表示向用户分配资源所花费的总时间或完成时间。

5.3 负载均衡策略正如上面分类中所

讨论的,几种技术被分为不同的类别,例如针对系统状态的静态和动态。动态算法使用系统状态和一些策略来执行任务 (Alakeel 2010; Daraghmi 和 Yuan 2015; Eager 等人1986; Kanakala 和 Reddy 2015; Mukhopadhyay 等人2010) ,如下所示:

·选择策略:此策略标识应从一个节点转移到另一节点的任务。它根据迁移所需的开销量、非本地系统调用的数量以及执行任务所需的时间来选择任务。 ·位置策略:该策略确定负载不足或空闲的计算节点,并将任务转移给它们进行处理。它根据可用的方法:探测、协商和随机,在确定任务迁移所需服务的可用性后选择目标节点。在随机方法中,位置策略随机选择目的地并传输任务。在探测方法中,节点探测其他系统节点以选择目的地。在协商方法中,节点相互协商以实现负载均衡。

·传输策略:该策略发现任务需要从本地节点传输到另一个本地/远程节点的情况。它由两种方法组成 所有当前任务和最后接收的任务 来识别要转移的任务。在最后接收方式中,所有传入的任务都进入转移策略,最后到达的任务将被转移。目前所有的任务方法都是基于规则的转移策略来决定任务需要转移 (任务迁移)还是本地处理 (任务重新调度),这取决于每个节点的工作负载。 ·信息策略:这是动态负载均衡的另一种策略,它将所有资源信息保留在系统中,其他策略进一步使用这些信息来做出决策。它决定了信息收集的时间。从节点收集信息的方法有代理、广播和集中轮询等。在广播方法中,所有节点广播它们的信息,其他节点可以访问这些信息。目前节点使用Agent方式来收集信息。各种信息政策包括需求驱动政策、周期性政策和状态变化驱动政策。

所有这些动态负载均衡策略都与进入系统的任务最初由传输策略处理有关。处理后,策略决定任务是否应转移到远程节点。对于需要转移的任务,位置策略会识别空闲或负载不足的目标节点。如果远程节点无法执行,则该任务将被放入队列中以在本地处理。两者均转

策略和位置策略从信息策略中收集所需的信息以做出决策。

6 现有负载均衡技术回顾

我们调查了有关现有负载均衡技术的文献,并对所选论文进行了彻底审查。根据系统状态,负载均衡技术主要分为静态负载均衡和动态负载均衡。

在这里,我们讨论现有的负载均衡技术:

6.1 静态负载均衡技术静态负载均衡技术不需要了

解系统的当前状态。它们只预先保留系统资源的知识,例如执行时间、内存、存储容量和节点的处理能力。静态负载均衡不允许在执行时分配资源。这些技术易于实现和执行,但对于资源较少的较小系统或网络非常有用。因为它们没有考虑系统的当前状态,所以这些技术对于执行分布式计算的计算机系统没有用处。它们也不允许在执行时检测连接的服务器计算机,这会导致资源分配不均匀。一些静态负载均衡技术包括循环、最小-最小、最大-最小、节流和先进先出。下面讨论一些静态技术。

用于负载均衡的动态交换算法允许在可用机器上分配有限负载 (Houle et al. 2002)。霍尔等人。(2002)考虑处理器上具有固定负载和同步通信的静态负载均衡。它仅使用本地计算在单端口令牌模型中以单位大小的作业(称为令牌)分配任务。它不允许在重新分配负载之前更改负载。它对于动态改变分布式环境中执行的负载和处理器没有用处。

静态负载均衡技术在最小化任务等待时间之前保持所有与任务相关的信息可用 (Kokilavani 等人, 2011)。它获取所有传入任务的列表并确定它们的执行时间。执行时间最短的任务将首先执行,执行时间最长的任务将最后处理。最短的作业将首先执行。然而,系统可能会面临某些任务的饥饿,因为它们不在执行队列中,因此需要更多的处理时间。基于静态任务的负载均衡允许任务以循环方式执行 (Pasha et al. 2014)。每个任务执行一个时间片,然后再次放入队列中执行另一个进程。遵循负载均衡技术,直到所有进程完成其任务。此技术对于所有到达的请求都具有相同类型的 Web 服务器非常有用。然而,这对于云环境来说并不好,因为云环境中的进程具有不同的配置。

6.2 动态负载均衡技术根据6.1节的讨论,静态技术

不需要了解系统的当前状态,因为这些技术不适合动态改变状态并一次需要当前资源状态的分布式计算系统间隔。因此,我们需要适合云环境的动态负载均衡技术。在本节中,我们讨论依赖于负载均衡器标准的各种动态负载均衡技术。根据我们的观察,我们将它们分类如下:

- 通用负载均衡 · 基于自然
- 现象的负载均衡 · 混合负载均衡

· 基于代理的负载均衡 · 基于任务的负载均衡 · 基于集群的负载均衡

6.2.1 一般负载均衡技术。在这里,我们讨论各种通用技术,例如虚拟机迁移和基于负载估计的算法(取决于各种负载均衡参数)。

Alakeel (2010)针对开发或研究负载均衡算法时需要考虑的问题编写了指南。作者讨论了动态负载均衡的不同组成部分,例如信息、传输和位置策略。作者涵盖了以下问题:负载评估、负载级别比较、性能指标、稳定性和节点之间的信息交换量。此外,作者还讨论了云负载均衡算法的局限性,得出的结论是当前的负载均衡策略在将作业分配给 VM 时不会保留有关 VM 先前状态的信息。负载均衡算法需要在每次新请求到达数据中心进行分配时运行。作者还利用CloudSim工具包用Java语言实现了循环算法。

肖等人提出了一种算法来实现两个主要目标:避免云环境中的过载和通过优化使用的服务器来实现绿色计算。(2013)。作者引入了偏度的概念来评估服务器的使用情况,并改善了多维约束下资源的整体使用情况。他们还提出了一种负载预测算法,有助于识别应用程序未来的资源消耗。当总体资源使用低于绿色计算阈值时,它会自动调用绿色计算算法来关闭一些未充分利用的服务器。作者根据从不同来源收集的各种数据跟踪,使用跟踪驱动模拟对该算法进行了评估。他们的结果表明,该算法改善了整体资源使用并最大限度地降低了功耗,但它可能导致某些资源过载。

Pop 等人提出了一种针对云环境中非周期任务的截止时间约束调度算法。(2015)。作者提出了一种方法,可以识别处理一组非周期性任务所需的数据资源数量。它将截止时间或执行时间和数据传输成本视为非周期性任务调度的主要约束。作者从大数据的角度挑战了传统的超级计算调度算法,并基于数学模型证明了一组来自不同来源的任务可以被视为一个任务。作者还证明,当所需资源数量超过数据中心的容量并且异构环境中的问题需要更多资源时,区域中心的任务迁移是最佳解决方案。结果将有助于优化数据中心的资源使用。

王等人。(2015)提出了一种框架设计来平衡名为 Swift 的云存储系统上的工作负载。作者表示,设计的框架能够识别集群中过载和负载不足的节点。他们设计了一种单独的负载均衡和资源分配算法,以在虚拟机上分配任务,这也有助于正确管理资源。该框架驻留在用户侧,不需要对用户操作系统和存储系统进行任何更改。它在监管范围内起作用,并且仅适用于较少数量的节点。但是,如果节点数量较多且具有多个范围且适用于不同的存储系统,则该方法不起作用。

需要使用负载均衡技术来估计任务的完成时间,并考虑虚拟机上的当前处理能力和作业大小。钱等人。(2016)提出了完成时间估计的两个因素:(1) 选定的虚拟机应提前完成工作,(2) 负载均衡算法必须估计排队作业和正在排队的作业的完成时间。接下来

进入队列。它会选择较早响应的VM进行作业分配并分配任务来执行。结果表明,它提高了响应时间和处理时间,但会导致功耗和碳排放增加。

Bala 和 Chana (2016)指出,为了在执行过程中主动管理云上的负载,应该使用机器学习方法在节点上进行预测。他们提出了一种负载预测模型来识别过载和负载不足的机器,以使用多个资源参数实现主动负载均衡。作者使用随机森林 (Cloudsim 工具环境中的一种机器学习方法)实现了该模型,通过考虑不同的资源使用参数来增强虚拟机中的负载均衡。所提出的模型仅在模拟环境中进行了测试;未来可以在实时环境中进行测试。此外,可以通过将容错技术合并到所提出的预测模型中来增强该模型。

内格利亚等人。(2016)提出了一种基于可再生能源供电的微型数据中心之间的负载均衡的平均场技术。该技术优先考虑由可再生能源供电的数据中心执行任务。所提出模型中的平均场技术有助于推断各种性能指标。

负载均衡技术用于最大限度地缩短用户请求的响应时间,并将它们统一分配给可用资源。 Ghoneem 和 Kulkarni (2017)修改了云分析模拟器上可用的主动虚拟机负载均衡技术。现有的虚拟机负载均衡技术在高峰时段的请求分配不均,导致用户的响应时间较长。他们使用预留表和分配表来维护有关请求的 VM 预留的信息。数据中心考虑负载均衡器提供的此信息,并将虚拟机分配给即将到来的用户请求。结果表明,即使在交通繁忙的情况下,所提出的技术也能为用户请求提供最短的响应时间。

陈等人。(2017)提出了一种相对于静态负载均衡方法的新颖负载均衡方法,主要考虑服务器处理能力和计算机负载,以最大限度地减少服务器上的负载。他们描述了现有算法中使用的各种负载均衡指标,并与提出的算法进行了比较。然而,他们没有考虑任务的迁移时间和响应时间,可以通过现有参数进行改进,以提高算法的负载均衡性。我们在表 3 中彻底分析并总结了一般的负载均衡技术。

6.2.2 自然现象 负载均衡技术的启发。表 4 总结了一些受自然现象启发的负载均衡技术,例如遗传算法、蜜蜂启发、ACO 和 PSO。

Babu 和 Krishna (2013)提出的受蜜蜂行为启发的负载均衡技术有助于将执行任务从过载的机器转移到负载不足的机器。它还平衡要在计算机上执行的任务的优先级,从而最大限度地减少任务的等待时间。它的工作原理是蜜蜂 (即任务)的概念,它向其他蜜蜂发出食物来源 (即负载不足的机器)的请求。它将任务的优先级作为主要的 QoS 参数。所提出的算法适用于在虚拟机上执行的独立任务,但不适用于相关任务。可以修改该算法以适用于相关任务。

可以包含的其他参数是执行时间和响应时间。

德法尔科等人。(2015a)提出了一种基于极值优化 (EO) 的技术,用于动态环境中并行执行用户请求的负载均衡。基于集合群体的并行EO方法考虑了适应度函数和目标节点的选择等因素,为动态方法提供了解决方案。在仿真过程中,将结果与基于 EO 的顺序算法的结果进行了比较 (De Falco 等人, 2014 年, 2015b)。变体 PEO-GS-D 基于

表 3. 通用负载均衡技术总结

参考	系统状态	技术	概念	优点	缺点
(王等人, 2015)	动态的	工作量 平衡和 资源 管理框架	实时虚拟机 移民	任务执行力 低	同质虚拟机, 独立任务
钱等人。 (2016)	动态的	迅速存储 使用作业完成时间估算方法	负载均衡 使用结尾的虚拟机 服务时间	减少响应时 间、减少处理时间	实际即时处理能力 计算较困难,功耗较多
巴拉和 查纳(2016)	动态的	使用机器学 习的预测负载均衡 方法	使用机器学习 识别过载和欠载的 机器	资源利用率高,迁 移开销低, 迁移次数减 少	未在真实云上进行测试
戈尼姆与库 尔卡尼(2017)	动态的	修改活性 VM负载均衡技术	使用预留 表统一分配请求	任务的最短 响应时间,负载均 衡 VM,提高弹性	单一数据中心统一 分配任务
陈等人。 (2017)	动态的	用于最 小化负载的新颖负 载平衡方法 在服务器上	动态附加方 法优于静态负载均衡	改进 资源利用, 完工时间,以及 服务质量	响应时间更长且考虑 的负载均衡参数 很少

表 4. 受自然现象启发的负载均衡技术

参考	系统状态	技术	概念	优点	缺点
巴布和 克里希纳 (2013)	动态的	蜜蜂行为 启 发负载均衡	利用蜜蜂的觅食行为	完工时间和响 应速度低 时间	不适用于相关任务
德法尔科等人。 (2015a)	动态的	基于极值优 化的负载均衡	动态并行任务执 行 环境	执行时间更短,任 务传输次数 更少,任务传输 次数更高 资源利用	不支持图优化和多目标 优化
巴布和 塞缪尔(2016)	动态的	增强的基于蜂群的 负载均衡	使用蜜蜂技术来 最大限度地 减少 资源 消耗和响应时间	响应时间短、高 资源使用,任务迁 移数量减少	可扩展性低、复杂性
(黛维和 乌达里亚拉吉 2016)	动态的	加权循环法	在考虑执行时 间的同时减少任务的 响应时间	响应时间短	同质环境执行

并行分支的解决方案给出了平均情况下的最佳结果。该算法不适用于应用图优化和多目标优化。

需要负载均衡技术来改善资源使用、最大限度地缩短云上任务的响应时间和完成时间。Babu 和 Samuel (2016)提出了一种将响应时间、QoS 和迁移数量考虑为负载均衡参数的技术。他们认为任务是蜜蜂,而未充分利用的虚拟机是食物来源。当虚拟机过载时,一些任务(低优先级的任务)会从一台虚拟机迁移到另一台虚拟机。该算法可以通过其他受自然启发的算法(例如 PSO 和 ACO)来增强。

Devi 和 Uthariaraj (2016)提出了一种改进的加权循环算法,通过考虑任务大小、VM 的容量以及多个任务的跨学科性质来提高任务执行的响应时间。该算法考虑所有 VM 的当前负载和执行任务的完成时间,并识别完成时间最短的机器。它在完成每个任务后运行负载均衡器以在节点之间分配负载。所提出的算法不在异构环境中执行,可以通过添加来自不同环境的机器来改进以获得一致的结果。我们分析了表 4 中受自然现象启发的负载均衡技术。

6.2.3 混合负载均衡技术。混合负载均衡技术的开发是为了通过保留静态和动态技术的特性和优点来克服静态和动态技术的缺点。混合技术可最大限度地缩短响应时间并提供高效的资源利用。我们在表 5 中总结了混合负载均衡技术。

王等人。(2013)提出了混合云的资源调度技术。他们考虑各种参数,例如执行时间、完成时间以及公共云和私有云的利用率。基于这些参数,他们提出了具有 OoS 满意度 (AsQ) 的自适应调度,以在私有云上优化资源分配。他们使用各种运行时估计和调度算法来确定私有云中的最佳资源分配。通过这个流程,他们减少了需要转移到公共云执行的任务。对于需要转移的任务,采用最小成本策略,尽量减少任务转移到公有云的比例。他们进行了许多实验,结果表明,与许多现有技术相比,任务的执行时间和完成时间被最小化。他们的技术有更好的 QoS。然而,该技术没有考虑其他重要参数,例如能源效率和运营成本,这些参数是修改将任务转移到公共云的模式所需要的。

Arab 和 Sharfi (2014)开发了一种用于负载均衡和资源发现单元的通信模型。每个节点都单独配备负载均衡和资源发现单元。负载均衡单元与其他节点中的同一单元交换消息,并获取有关可用资源的足够信息。这些信息有助于资源发现单位更准确地做出决策。根据所提出的模型,负载均衡器负责使用 Dodonov 和 De Mello (2010) 描述的各种技术来提取处理器的状态。它提供了可扩展性和较短的响应时间,但会导致资源发现时间增加。

由按需调度、查询和迁移任务 (QMT) 和分阶段任务迁移 (STM) 组成的混合负载均衡方案有助于更有效地管理节点上的负载 (Liu et al. 2015)。每当一个节点检测到负载较重时, QMT 就会通过将最后传入的节点转移到另一个负载较低的节点来平衡负载。 QMT 和 SMT 都适用于依赖和独立任务,但它们都面临着传输和调度时间长的的问题,而这些时间可以在将来最小化。

表 5. 混合负载均衡技术总结

参考	系统状态	技术	概念	优点	缺点
王等人。 (2013)	动态的	并行任务的 自适应调度技 术	用于任务转移的最 小成本策略以改善资 源	更高的资源利用 率、最小化的任 务迁移、更好的 QoS	没有考虑运 营成本、能源效率等重要 参数。
阿拉伯和 谢里菲(2014)	动态的	资源发现和 负载均衡组合	使用 提高可扩展 展性 沟通 资源和负载之间 的平衡。	高度可扩展， 不需要其他节点信 息	资源发现时间长， 策略传播
曹等人。 (2015)	动态的	动态蚁群算法与粒子群优化算 法相结合	结合 ACO 和 PSO 来改善资源	增加资源使用，减少计算 时间	计算成本高，同构服务器 支持
刘等人。 (2015)	动态的	混合负载均衡和 任务调度	使用 基于主节点的混合 负载均衡实现从节 点的负载均衡	独立和依赖任务调度， 响应时间更低	高传输和调度 时间
那霸和 奥斯曼 (2016)	动态的	代理和负载均衡技术 的结合	结合负载均衡 和经纪人技术 来减少响应 时间	处理时间短、响应低 时间	低性能、高执行力 时间
(陈等人， 2017)	动态的	云负载 均衡 (CLB)技术	架构至 克服服务器 响应失败和负载均衡 器来监控优先级、计算 能力	高度可扩展	高响应 时间

曹等人。 (2015)通过结合 PSO 和 ACO 的特征,提出了一种混合负载均衡技术。每次请求到达时,它都会检查每个服务器上的负载并确定最大可用内存。所提出的技术根据服务器上的负载历史来预测即将到来的负载,由于定期分析服务器上的可用内存,因此在内存和时间的情况下会导致更高的计算成本。根据内存的可用性和任务的大小,它在调度之前接受或拒绝传入的请求。结果,接受的请求轮流执行,这使所有服务器保持忙碌。它可以提高任务的资源利用率和计算时间。该技术通过根据有关任务的历史信息预测即将到来的请求的工作负载来符合动态执行环境。预拒绝步骤拒绝不满足预调度标准的请求,从而最大限度地减少计算时间。

作者指出,所提出的技术比传统的 ACO 技术快得多,但它仅适用于同构服务器,并且没有考虑客户端和服务提供商的成本。

Naha 和 Othman (2016)通过将循环和节流负载平衡技术与性能优化的服务代理和服务邻近代理算法相结合,实现了一种混合算法来优化系统性能。他们提出了三种服务代理算法和一种负载平衡算法。作者将它们描述为 Load

表 6. 基于代理的负载均衡技术总结

参考	系统状态	技术	概念	优点	缺点
陈等人。 (2013)	动态的	云端新颖的紧急任务分配	采用公平竞争和动态调整原则的云端紧急任务分配	高效的资源配置	处理时间越长, 传输时间
塔斯奎尔 (2015)	动态的	基于多代理的负载均衡	在多云环境中使用多个代理进行资源配置和监控	提供极高的弹性设施, 使用多云 资源	未实现, 不考虑QoS
加西亚和纳法拉特 (2015)	动态的	基于代理的负载均衡	使用代理迁移虚拟机	异质虚拟机和服务器	迁移开销高,可扩展性低
克沙瓦迪和法吉(2016)	动态的	基于多代理的负载均衡架构	通过多个代理最大限度地利用资源	响应时间短、完工时间缩短、改进 资源利用	DCM 代理需要父消息来销毁自己, 没有自我销毁计时器

Aware (LA) 和 Cost Aware (CA) 算法可提高资源利用率。 LA 算法提供较短的处理时间,但产生较高的成本,而 CA 算法则降低成本。服务代理算法根据用户需求选择服务器,这可能会导致处理时间或成本的增加。当CA用户请求服务器上的com时,他们使高速 (高成本)服务器保持空闲,这降低了资源使用并增加了用户请求的计算时间。服务邻近代理算法选择距离客户所在区域最近的数据中心。节流负载均衡器维护所有可用虚拟机的表。作者将所有提到的服务代理和负载均衡算法组合在一起来满足用户请求;结果表明处理时间和响应时间都减少了。但请求的执行时间较长,系统性能也有待提高。

需要开发有效的云负载均衡 (CLB)架构来克服用户请求较多时服务器响应失败的问题。陈等人。 (2017) 开发了考虑计算机负载和服务器处理的架构,以最大限度地减少服务器问题以处理更多计算请求。他们还物理和虚拟 Web 服务器提供了一种负载均衡技术,以保留有关服务器负载、优先级和计算能力的信息。该架构提供了高度可扩展的性能,但会导致响应时间增加。我们在表 5 中全面总结了混合负载均衡技术。

6.2.4 基于代理的负载均衡技术。在这里,我们调查了各种基于代理的负载均衡技术。代理用于云资源发现、协商、组合和管理。代理自动连续工作以满足设计目标。

多个代理可以协同工作,以满足用户的 QoS 要求和资源使用。在表6 中,我们总结了这些技术。

陈等人。 (2013)开发了一种云紧急任务分配技术。为了实现负载均衡,作者采用公平竞争原则和动态调整原则来适应到达的任务。该技术使用轮盘赌机制,其中投标人参与投标过程以获得分配给资源的任务并使用缓冲区

120:20

P. 库马尔和 R. 库马尔

强化绩效池机制的办法。实验结果表明

这些技术有效地分配资源,但增加了处理和传输时间。

应用程序感知、基于多代理的负载平衡架构有助于自动配置资源。Tasquier (2015)指出,开发的架构使用三种不同的

代理(执行程序代理、配置程序代理和监视代理)分别负责表示正在运行的应用程序、缩放资源以及监视资源的过载/欠载情况。所提出的算法还提供了

用于检查云和资源弹性当前状态的工具。然而作者并没有

在云环境中实现算法来测试有效性,不考虑QoS。

Garcia 和 Nafarrate (2015)提出了一种基于协作代理的负载平衡技术,该技术

使用实时虚拟机迁移概念来处理异构服务器之间的负载。他们还提出了一种基于代理的负载平衡机制,该机制由识别虚拟机的程序组成

需要与目的地一起迁移的内容、虚拟机迁移策略、虚拟机接受策略以及为虚拟机选择初始主机的负载平衡程序。

作者指出

与提高任务响应时间和资源使用的集中式技术相比,该算法在负载平衡方面表现更好,但它增加了迁移开销。

基于多代理的负载平衡架构有助于最大限度地利用资源(Keshvadi 和

法吉赫2016)。它执行发送者发起和接收者发起的技术以减少

任务的等待时间并保证SLA。该模型由以下代理组成: 虚拟代理

机器监控代理(VMM代理)、数据中心监控代理(DcM代理)和协商器

蚂蚁代理(NA代理)。VMM代理支持系统中的所有VM并保存有关信息

VM使用内存、CPU和带宽来监控负载。DCM代理执行信息

策略使用VMM代理提供的信息,并根据不同的特征对VM进行分类。它还启动NA代理,这些代理会移动到其他数据中心以了解

那里可用的虚拟机的状态。仿真结果表明该方法更加高效并且改进

响应时间和完工时间。我们在表6中分析了基于代理的负载平衡技术。

6.2.5 基于任务的负载平衡算法。在这里,我们讨论基于任务的文献

负载平衡技术。大多数现有的负载均衡方案更倾向于通过以下方式迁移虚拟机:

将过载的虚拟机从一台物理机转移到另一台物理机。它使

资源分配灵活,但需要更多的时间和成本,因为我们需要转移整个虚拟机

机器而不是由于机器超载而转移任务。表7

总结了现有的基于任务的负载平衡技术。

拉梅扎尼等人。(2014)开发了一种基于任务的系统负载平衡(TBSLB)方法。

该方法基于PSO技术来降低虚拟机的迁移成本。他们还有

开发了转移任务的优化模型。他们通过扩展嵌入Jswarm包的CloudSim来评估所提出的模型(Calheiros等人,2011)。所提出的机制优于当时现有的负载平衡技术,因为它最大限度地减少了停机时间

VM、内存使用和成本。他们研究了参数传输时间、内存停机时间和使用成本。稍后,可以针对多目标PSO和其他负载均衡进行改进

参数。

吴等人。(2016)提出了一种基于遗传和蚁群的任务调度算法。他们

结合了遗传和ACO技术的特点。遗传算法搜索

在早期阶段可用资源,然后蚁群算法为任务选择最佳资源。单独来看,遗传算法响应时间较长,效率较低,

和冗余。由于蚁群算法在资源搜索阶段也效果不佳

缺乏信息素。通过结合两种算法的特征,混合方法改进了

虚拟机之间的负载平衡并提高效率。

表 7. 基于任务的负载平衡技术总结

参考文献	系统状态	Ramezani	技术	概念	优点	缺点
Dynamic 等人。 (2014)			基于任务的粒子群优化 系统负载均衡方法	实时虚拟机迁移 低任务执行	同质化	VM,独立任务
吴等人。 (2016)	动态的		任务调度的遗传-蚁群混合算法	结合遗传和蚁群优化技术的特点,对可用资源进行遗传搜索,蚁群选择最优解执行。	改进负载均衡、提高效率、最小化执行时间	未在实际云环境中测试
沉等人。 (2016)	动态的		考虑网络拓扑和传输调度中的时间	实现任务调度 MapReduce 最大限度地减少数据传输时间和成本	改进集群使用,最大限度地缩短作业完成时间	无带宽预留,模型未在不同网络条件下进行评估
埃尔穆吉等人。 (2017)	动态	一种基于最短作业和动态任务量子循环的混合任务调度算法		将短任务和长任务存储在单独的就绪队列中并分别执行以最大程度地减少饥饿	最大限度地减少任务的饥饿和等待时间,提高响应时间和周转时间	任务量程为效果较差
辛等人。 (2017)	动态的		用于并行任务的经济的多个调度程序	多个调度程序执行并行任务和加权 机器分配方法	提高资源利用率,最大限度地减少任务权重和执行时间	参数评估不是最优的

沉等人。 (2016)提出了一种用于 MapReduce 的网络感知任务放置技术,以减少作业完成时间、总体传输时间和数据成本。他们指出,任务主要面临以下问题:第一,资源可用性由于随时间的释放和访问而动态变化;其次,减少任务的数据获取时间取决于任务的大小和位置;第三,路径上的负载也会对数据访问延迟产生相关影响。为了最大限度地减少数据访问延迟,在任务调度决策期间应考虑路径上的负载。结果表明,它最大限度地缩短了任务的完成时间并增加了资源的使用。

多调度器架构有助于最大限度地减少执行大规模并行任务及其执行成本的问题。辛等人。 (2017)提出了一种基于加权随机调度的调度方法,以最小化任务之间的资源竞争和高设备负载。将考虑执行时间、成本和通信延迟等参数来为任务分配权重。成本较低的机器将获得较高的成本,并且有更多的机会被分配执行任务。作者使用 WorkFlow 生成器对 MATLAB2012b 进行了实验,以生成所需的数据集。他们在数据集上进行了实验,这些数据集包含一个时间间隔内较大的任务集,以及执行时间、成本和传输延迟。他们还考虑了任务结构、任务到达时间、设备依赖性和特定设备集测试。结果表明,多个调度器改进了任务等参数

设备和执行成本的竞争。然而,它尚未评估参数的最佳值,可以进一步改进。

用于任务分配的选择具有最小负载的虚拟机 (SVLL) 负载均衡模型提高了云系统的性能 (Aladwani 2017)。所提出的模型计算每个VM的负载,并根据VM的负载而不是分配给VM的任务数量来分配要执行的任务。Aladwani (2017)将 SVLL 技术与其他任务调度技术 (例如先来先服务和最短作业优先)一起实施,从而改善了任务的总等待时间和总完成时间。该算法与基本调度算法一起实现,与单独使用基本算法相比,可以给出更好的结果。

埃尔穆吉等人。(2017)开发了一种任务负载平衡技术,结合了最短作业优先和循环调度技术的特点。它将短任务和长任务存储在两个独立的就绪队列中,并使用动态任务量来平衡任务之间的等待时间。作者将吞吐量和饥饿问题放在一起考虑。他们在 CloudSim 模拟器上评估了该算法;结果表明,周转时间、等待时间和响应时间都被最小化。它还最大限度地减少了长期任务匮乏的情况。然而,任务量在平衡任务之间并不是很有效,未来可以改进以改善等待时间和饥饿问题。我们在表 7 中分析了基于任务的负载平衡技术。

6.2.6 基于集群的负载平衡技术。在异构云环境中,各种资源被放置在不同的数据中心,并根据服务器性能、存储容量等参数划分为不同的集群。在本节中,我们讨论一些基于集群的技术并将其总结在表 8 中。

基于预测的自适应负载平衡技术有助于改善数据中心资源的使用。毛等人。(2013)指出,所提出的技术可以预测集群中的负载,如果工作负载低于最小阈值,则相应地回调集群中的资源,并在工作负载高于阈值时添加新的虚拟机。作者在 Cloudsim 模拟器上实现了该算法;结果表明,它提高了资源利用率,并最大限度地缩短了任务的响应时间。

Daraghmi 和 Yuan (2015)为开发的控制器提出了一种负载平衡机制,以克服大型数据中心集中式控制器的问题限制。它将整个网络划分为多个区域,并在每个区域分配一个控制器,通过分析开发的控制器中的不平衡负载问题来重新路由流量,以进行流量控制和网络管理。在所提出的技术中,当发生不平衡负载时,它将一部分负载迁移到控制器以动态管理负载。作者为 LBDC 技术开发了多种解决方案,包括贪婪方法和一种分布式贪婪方法来管理流量,并将结果与现有开发的控制器进行了比较。

赵等人。(2016)提出了一种基于贝叶斯和聚类 (LB-BC) 的负载平衡启发式技术,以克服现有负载平衡方法的复杂性。该方法基于贝叶斯定理 (Agostini 1995; Pawlak 2001)并实现了长期负载平衡。

它计算物理主机的后验概率,并将其与聚类思想相结合以选择最佳主机。它考虑以下参数:请求的任务数量、标准偏差和负载平衡效果。后来,将该方法与动态负载平衡进行了比较,随着时间的增加,动态负载平衡导致标准偏差最小。所提出的方法仅适用于局域网,但可以针对广域网和实时环境进行增强。

Kang 和 Choo (2016)提出了一种基于集群的作业调度技术,以改善云间通信,以实现动态和实时多媒体流中的负载平衡。所提出的技术有一个两步过程。首先,它创建用于监控活动、管理

表 8. 基于集群的负载均衡技术总结

参考	系统状态	技术	概念	优点	缺点
给我和元(2015)	动态	基于小型世界的覆盖网络,用于动态负载均衡改进	分散式大型数据中心基于贪婪方法开发的控制器	改善资源利用,提高数据中心性能	高功耗
姜与秋(2016)	动态的	基于集群的作业调度技术	改善大规模云间通信,以实现动态实时多媒体流中的负载均衡	更好的响应时间 由于拥塞而丢包	
赵等人。(2016)	动态的	基于贝叶斯和聚类的负载均衡 (LB-BC)	物理主机的后验概率及其与聚类概念的结合	最小标准偏差和响应时间	仅适用于 LAN,不适用于实时环境
汉和克罗诺普洛斯(2017)	动态的	分布式模型用于改善负载均衡的自调度技术	一类分布式自调度方案,提高负载均衡和可扩展性	提高了可扩展性,提高了整体性能,减少通信开销	不支持并行执行

平台复杂性,并在定期向所有连接的邻居服务器广播 hello 数据包的基础上满足主机的需求和令人满意的 QoS。其次,如果作业的等待时间超过队列中的 5 个任务,则决定转移作业请求。

该技术的性能与蚁群、WCAP 和 HFA 相比,具有更高的作业调度率,并且提供了更好的响应时间。此外,该技术可以针对中间节点由于资源限制而拥塞的实时环境进行改进,并且可以使用通信作业而不是计算作业来减少由于拥塞而导致的数据丢失。

泽格拉里等人。(2016)提出了一种基于集群的任务负载均衡技术来克服节点上的负载分配问题。它结合了遗传算法和库恩算法的概念,通过将任务分组并分布在协作节点中来开发任务分配策略。所提出的算法提供了更好的响应时间和数据中心节点之间的任务分配。

Han 和 Chronopoulos (2017)开发了一种用于自调度方案的分层分布式模型,以提高云系统的负载均衡和可扩展性。该模型能够在同构和异构环境中执行。作者使用四种不同的计算应用程序在大规模集群中应用了这些方案;结果表明,可扩展性得到提高,整体性能得到提高,通信开销也得到降低。

未来,该算法可以针对大规模集群和具有依赖性的循环进行测试。我们在表 8 中分析了基于集群的负载均衡技术。

6.2.7 负载均衡和资源调度工具。在本节中,我们讨论一些可用的云模拟器 (Calheiros et al. 2011; Garg and Buyya 2011; Gupta et al. 2011; Jararweh

120:24

P. 库马尔和 R. 库马尔

等人。2013年;努涅斯等人。2011;奥斯特曼等人。2010;维克勒马辛哈等人。2010)。云模拟器是一个框架,当由于成本因素而无法开发用于评估这些因素的真实云架构时,云模拟器提供虚拟环境来测试性能、部署模型、资源分配和资源负载平衡。在部署到实际硬件上之前,模拟器在测试和验证负载平衡、调度算法方面发挥着重要作用。所有模拟器都有一些独特的功能,用于评估不同的参数性能。表9显示了可用模拟器的功能、可用性、使用的编程语言、优点和缺点。

7 讨论

我们根据这些技术在云计算环境中使用的各种指标对研究文章进行了比较分析,如表10和图5所示。表10显示了各种研究论文如何考虑不同的 QoS 指标。综合回顾各种负载均衡技术后,可以说不同的技术考虑了不同的评估指标。一些论文考虑了单一目标,而另一些论文则考虑了指标的多个目标。图5显示了不同文章考虑的负载平衡指标的百分比。图6显示了不同研究论文中使用的评估工具。

我们还根据模拟和可用结果分析了各种研究论文。

表10和图5显示,大多数用户都关注即将执行的任务的响应时间,其次是资源使用、完工时间和迁移时间。研究人员专注于最小化响应时间以维持 SLA 并增加响应时间。

Devi 和 Uthariaraj (2016)、Babu 和 Krishna (2013)以及 De Falco 等人。(2015a)认为响应时间是评估算法的主要参数。

8 研究趋势

在云负载均衡方面,仍然存在许多问题和挑战需要未来讨论和重新解决。从文献回顾中,我们发现了云未来需要改进的一些方向。为了维持云性能,需要考虑一些未来的方向:服务质量 (QoS)、服务水平协议、资源配置和负载平衡等问题。云服务提供商需要大量资源来维护 QoS 和 SLA。SLA 是根据 QoS 规则设计和部署的,如果有任何违反 SLA 的情况,服务提供商必须支付罚款。自动资源配置最大限度地减少了用户和服务提供商的交互。为了维护 SLA 和 QoS,需要负载平衡以正确使用所配置的资源。负载均衡器有助于以高吞吐量保持最少的资源量。已经开发出许多考虑不同指标的负载平衡技术,例如性能、响应时间、执行时间、任务迁移时间和资源使用。没有技术考虑到所有负载平衡参数,从而提高数据中心的整体性能。

云负载均衡存在重要的未解决问题和挑战,如下所示:

- 需要为工作负载正确分配资源以提高云性能。 · 功能需求 (SLA)和非功能需求 (QoS)都应该满足保持。
- 管理即时发生的大量用户请求以及管理当前执行任务的 SLA。 · 在与不同服务提供商执行的情况下维持应用程序成本和异构环境。

图 1	图 1 展示了云计算中负载均衡技术的架构。该架构由多个组件组成，包括客户端、负载均衡器、服务器和数据库。客户端通过负载均衡器连接到服务器，服务器通过数据库进行数据交互。
图 2	图 2 展示了云计算中负载均衡技术的性能指标。该指标包括吞吐量、延迟、可用性和可扩展性。吞吐量是指单位时间内处理的请求数量，延迟是指请求从发送到接收的时间，可用性是指系统正常运行时间，可扩展性是指系统处理负载增长的能力。
图 3	图 3 展示了云计算中负载均衡技术的部署模式。该模式包括集中式部署、分布式部署和混合部署。集中式部署是指所有负载均衡器都连接到同一个中央服务器，分布式部署是指负载均衡器分布在不同的服务器上，混合部署是指集中式和分布式部署的结合。
图 4	图 4 展示了云计算中负载均衡技术的配置参数。该参数包括负载均衡器的地址、端口、权重和超时时间。负载均衡器的地址是指负载均衡器的 IP 地址，端口是指负载均衡器监听的端口，权重是指负载均衡器分配给服务器的权重，超时时间是指负载均衡器等待服务器的响应时间。
图 5	图 5 展示了云计算中负载均衡技术的故障处理机制。该机制包括故障检测、故障转移和故障恢复。故障检测是指负载均衡器定期检查服务器的状态，故障转移是指当服务器发生故障时，负载均衡器将请求转移到其他健康的服务器，故障恢复是指当故障服务器恢复后，负载均衡器将其重新加入到负载均衡池中。
图 6	图 6 展示了云计算中负载均衡技术的性能测试结果。该结果包括吞吐量、延迟、可用性和可扩展性的测试数据。吞吐量测试结果表明，负载均衡器能够处理大量的请求，延迟测试结果表明，负载均衡器的延迟非常低，可用性测试结果表明，负载均衡器的可用性非常高，可扩展性测试结果表明，负载均衡器能够处理负载的快速增长。
图 7	图 7 展示了云计算中负载均衡技术的未来研究方向。该方向包括智能负载均衡、边缘计算和云原生负载均衡。智能负载均衡是指利用人工智能技术优化负载均衡策略，边缘计算是指将负载均衡器部署在网络的边缘，云原生负载均衡是指专门为云原生应用设计的负载均衡器。

	<div><div></div><div></div><div></div></div>
<div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div></div>
<div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div></div>
<div><div><div></div><div></div></div><div></div></div>	<div><div></div><div></div><div></div><div></div></div>
<div><div><div></div><div></div></div><div></div></div>	<div><div></div><div></div><div></div></div>
<div><div><div></div><div></div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
	<div><div></div><div></div><div></div></div>

1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

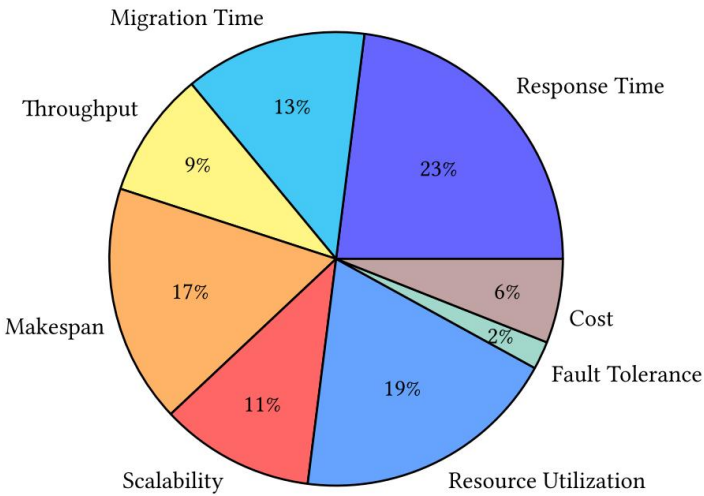


图 5. 调查技术中负载均衡指标的百分比。

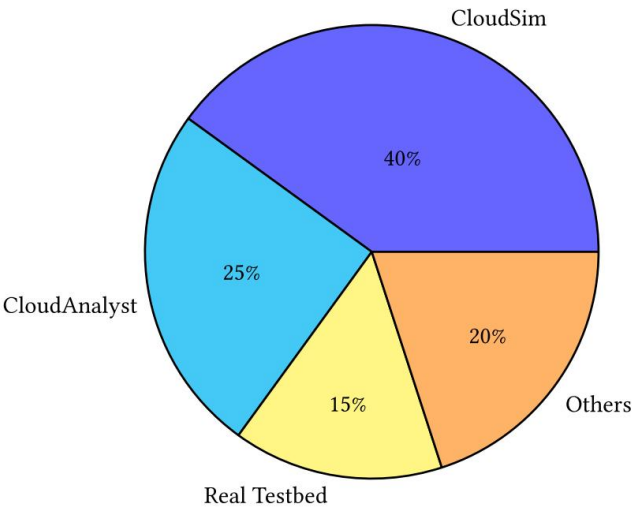


图 6. 调查论文中使用的评估工具/技术。

· 大量服务提供商正在建立数据中心。如何根据自己的需求寻找完美的服务提供商,对用户来说是一个更大的挑战。 · 随着云提供商的不断增长,可能会出现需要将工作负载转移到另一个云提供商的情况,并且由于不同的数据和服务策略可能会成为一项挑战。

· 当工作负载需要转移到另一个云时,数据锁定也可能是一个问题
· 提供商需要一些政策来解决此类问题。 · 通过组件级测试和基于检查点的方法
提高云系统的可靠性。 · 由于云服务和数据中心的需求不断增加,电力消耗

也在增加;最小化功耗是一个重要问题。

- 每天都会从不同来源生成大量数据,例如银行、社交媒体网站和电子商务,需要高质量的存储系统和一流的技术以便于检索和分析。
- 异构云环境中资源和应用程序的管理是一个非常复杂的任务。
- 设计良好的资源分配机制对于服务提供商来说是一个重要问题,不断提高资源利用水平。

9 结论

虚拟机上任务的负载均衡是云计算中的一个主要挑战,引起了研究人员的极大关注。本文代表了对问题和问题的最先进的回顾

负载均衡的挑战。根据这项研究,我们对考虑不同指标的多种负载均衡技术进行了广泛的调查。根据我们的知识,我们有

将这些负载均衡技术分为几类:一般负载均衡技术、基于自然现象的负载均衡、基于任务的负载均衡和基于代理的负载均衡

负载均衡。我们讨论了每个类别的优点、缺点、概念和

这些技术的挑战。有许多负载均衡算法可供使用,它们结合了大多数负载均衡指标,并提供更高的资源利用率和更少的响应时间。然而,

未来需要改进技术以提高系统的性能。

随着系统性能和资源利用率的提高,负载均衡技术

还强调绿色计算、节能和任务负载管理,这需要开发新的算法。因此,有必要对新开发的和

最近在模拟器上基于各种负载均衡指标的不同类别的负载均衡技术,以在实际部署之前检查这些算法的有效性

云环境。

这项研究将有助于研究人员确定负载中的研究问题平衡字段,并将提供可用负载均衡技术的摘要。

参考

沙菲·穆罕默德·阿卜杜勒哈米德、阿卜杜勒·拉蒂夫、穆罕默德·沙菲和穆罕默德·巴克里·巴希尔。2014。调度技巧
按需网格即服务云:回顾。理论与应用信息技术杂志 63, 1 (2014),
10-20。

GD阿戈斯蒂尼。1995。基于贝叶斯定理的多维展开方法。核研究所和方法

物理研究 A 362,2-3 (1995), 487-498。DOI:[https://doi.org/10.1016/0168-9002\(95\)00274-X](https://doi.org/10.1016/0168-9002(95)00274-X)

伊什法克·艾哈迈德和阿里夫·加富尔。1991。大规模并行多计算机系统的半分布式负载均衡。IEEE

软件工程汇刊 17, 10 (1991), 987-1004。DOI:<https://doi.org/10.1109/32.99188>

塔哈尼·阿拉德瓦尼。2017。选择负载最小的虚拟机对云计算任务调度算法的影响。第二届大数据、云和应用国际会议 (BDCA 17) 论文集。
ACM,纽约,

纽约州,第13条,第7页。DOI:<https://doi.org/10.1145/3090354.3090367>

阿里·穆罕默德·阿拉克尔。2010。分布式计算机系统动态负载均衡指南。国际期刊
计算机科学与信息安全 10, 6 (2010), 153-160。

穆罕默德·诺鲁兹·阿拉伯和莫森·谢里菲。2014。资源发现和负载之间的通信模型

平衡计算环境中的单元。超级计算杂志 68, 3 (2014), 1538-1555。DOI:<https://doi.org/10.1007/s11227-014-1124-y>

迪尼什·巴布 (Dhinesh Babu) 和文卡塔·克里希纳 (Venkata Krishna)。2013年。蜜蜂行为启发了云计算环境中任务的负载均衡。应用软计
算杂志 13, 5 (2013), 2292-2303。DOI:<https://doi.org/10.1016/j.asoc.2013.01.025>

雷梅什·巴布和菲利普·塞缪尔。2016。增强的蜂群算法,可在云中实现高效的负载均衡和调度。仿生计算和应用的创新 4 (2016), 135-142。
DOI:https://doi.org/10.1007/978-3-319-28031-8_6

安朱·巴拉和因德维尔·查纳。2016年。通过虚拟机迁移实现基于预测的主动负载均衡方法。工程师-
与计算机一起学习 32, 4 (2016), 1-12。DOI:<https://doi.org/10.1007/s00366-016-0434-5>

120:30

P. 库马尔和 R. 库马尔

Sourav Banerjee, Mainak Adhikari, Sukhendu Kar 和 Utpal Biswas. 2015. 开发和分析一种新的云分配策略,以提高云中的服务质量。阿拉伯科学与工程杂志 40, 5 (2015), 1409–1425。

DOI:<https://doi.org/10.1007/s13369-015-1626-9>

比巴尔·贝尼法和德杰。2017. 基于高效局部性和副本感知调度 (ELRAS) 策略的异构集群 MapReduce 性能改进。无线个人通信 95, 3 (2017), 2709–2733。

DOI:<https://doi.org/10.1007/s11277-017-3953-5>

阿迪亚·巴德瓦吉和查拉·拉玛·克里希纳。2018. 云计算中虚拟机迁移的高效多级带宽分配技术。智能与模糊系统杂志 36 (2018), 1–14。 DOI: <https://doi.org/10.3233/JIFS-169819>

Kyoungsoo Bok, Jaemin Hwang, Jongtae Lim, Yeonwoo Kim 和 Jaesoo Yoo. 2016. 一种用于处理大型多媒体数据的高效 MapReduce 调度方案。多媒体工具和应用程序 76, 16 (2016), 17273–17296。 DOI:<https://doi.org/10.1007/s11042-016-4026-6> 伊芙琳·布朗。2011 年。NIST 云计算定义的最终版本发布。

2017 年 11 月 10 日检索自 <https://www.bluepiit.com/blog/different-types-of-cloud-computing-service-models/>。

Rajkumar Buyya, James Broberg 和 Andrzej M. Goscinski. 2010. 云计算:原理和范式。卷。87. 约翰威利父子公司。

Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, César AF De Rose 和 Rajkumar Buyya. 2011. CloudSim: 用于云计算环境建模和仿真以及资源配置算法评估的工具包。

软件: 实践与经验 41, 1 (2011), 23–50。 DOI:<https://doi.org/10.1002/spe.995> 戈兰·坎德尔利克。

2013. 我能用它做什么: 云服务。2018 年 2 月 20 日检索自 <https://www.globaldots.com/cloud-computing-types-of-cloud/>。

耶加内·查班德和尼玛·贾法里·纳维米普尔。2016. 在线知识共享机制: 对最先进文献的系统回顾和对未来研究的建议。信息系统前沿 18, 6 (2016), 1131–1151。

DOI:<https://doi.org/10.1007/s10796-016-9628-z>

陈超、朱晓敏、鲍卫东、陈立东和光梦辛。2013. 基于代理的云紧急任务分配算法。第 10 届 IEEE 高性能计算和通信国际会议及 IEEE 嵌入式和普通计算国际会议 (HPCC_EUC 13) 论文集。IEEE, 1490–1497。

陈尚良、陈云耀、郭双红。2017. CLB: 一种新颖的云服务负载均衡架构和算法。计算机与电气工程 58 (2017), 154–160。 DOI:<https://doi.org/10.1016/j.compeleceng.2016.01.029>

Nguyen Khac Chien, Nguyen Hong Son 和 Ho Dac Loc. 2016. 云计算中基于估计服务完成时间的负载均衡算法。第 18 届 IEEE 国际先进通信技术会议 (ICACT 16) 论文集。IEEE, 228–233。

马丁·奇雷吉和尼玛·贾法里·纳维米普尔。2016. 一种在云环境中使用意见领袖实体的推荐并消除巨魔实体影响的信任和声誉评估的新方法。计算机在人类行为中的应用 60 (2016), 280–292。 DOI:<https://doi.org/10.1016/j.chb.2016.02.029>

Keng-Mao Cho, Pang-Wei Tsai, Chun-Wei Tsai 和 Chu-Sing Yang. 2015. 云计算中负载均衡虚拟机调度的混合元启发式算法。神经计算和应用 26, 6 (2015), 1297–1309。

DOI:<https://doi.org/10.1007/s00521-014-1804-9>

比亚吉奥·科森扎、根纳罗·科尔达斯科、罗萨里奥·德基亚拉和维托里奥·斯卡拉诺。2011. 基于并行代理的模拟的分布式负载均衡。第 19 届 Euromicro 国际并行、分布式和基于网络的处理会议 (PDP 11) 的会议记录。IEEE, 62–69。

埃曼·亚西尔·达拉格米 (Eman Yasser Daraghmi) 和袁贤明 (Syan-ming Yuan)。2015. 基于小世界的覆盖网络, 用于改进动态负载均衡。系统与软件杂志 107 (2015), 187–203。 DOI:<https://doi.org/10.1016/j.jss.2015.06.001> 苏曼·达斯、哈里什·维斯瓦纳坦和吉·里滕

豪斯。2003 年。通过分组数据系统中的协调调度实现动态负载均衡。IEEE 计算机与通信第 23 届年度联合会议论文集, IEEE 学会, 卷。1. IEEE, 786–796。

库西克·达斯卡普塔、布罗托蒂·曼达尔、帕拉玛莎·杜塔、乔茨纳·库马尔·曼达尔和桑塔努大坝。2013. 基于遗传算法 (GA) 的云计算负载均衡策略。Procedia 技术 10 (2013), 340–347。 DOI:<https://doi.org/10.1016/j.protcy.2013.12.369> 奇特拉·德维 (Chitra Devi) 和莱蒙德·乌塔里亚拉 (Rhymend Uthariaraj)。

2016. 云计算环境中的负载均衡, 使用改进的加权循环算法来处理非抢占式相关任务。《科学世界杂志》2016 年 (2016), 1–14。 DOI:<https://doi.org/10.1155/2016/3896065>

巴卡特·迪隆。2015 年 (2018 年 2 月 23 日访问)。不同类型的云计算服务模型。2018 年 2 月 23 日检索自 <https://www.nist.gov/news-events/news/2011/10/final-version-nist-cloud-computing-definition-published/>。

叶夫格尼·多多诺夫和罗德里戈·费尔南德斯·德梅洛。2010.一种基于通信事件预测的分布式应用程序调度的新方法。未来一代计算机系统 26, 5 (2010), 740–752. DOI:<https://doi.org/10.1016/j.future.2009.05.004>施里达尔·多马纳尔和拉姆·莫哈纳·雷迪。2013.使用改进的节流算法实现云计算中的负载均衡。

IEEE 国际新兴市场云计算会议 (CEEM 13) 会议记录。IEEE,1-5。

DOI:<https://doi.org/10.1109/CEEM.2013.6684434>

德里克·伊格·爱德华·拉佐夫斯卡和约翰·扎霍尔詹。1986.同质分布式系统中的自适应负载均衡。IEEE 软件工程汇刊 26, 5 (1986), 662–675. DOI:<https://doi.org/10.1016/j.future.2009.05.004>萨米尔·埃尔莫吉·沙亨达·萨尔汉和马纳尔·琼迪。2017.一种新颖的混合最短作业优先和循环与动态可变量子时间任务调度技术。云计算杂志 6, 1 (2017), 12. DOI:<https://doi.org/10.1186/s13677-017-0085-0>

伊万诺·德·法尔科·埃里克·拉斯科夫斯基、理查德·奥莱尼克、翁贝托·斯卡弗里、埃内斯托·塔伦蒂诺和马雷克·图德鲁伊。2014.分布式程序负载均衡中引导状态变化的极值优化。第 22 届 Euromicro 国际并行、分布式和网络处理会议论文集。IEEE,228-231. DOI:<https://doi.org/10.1109/PDP.2014.56>

1109/PDP.2014.56

伊万诺·德·法尔科·埃里克·拉斯科夫斯基、理查德·奥莱尼克、翁贝托·斯卡弗里、埃内斯托·塔伦蒂诺和马雷克·图德鲁伊。2015a.极值优化应用于分布式程序执行中的负载均衡。应用软计算杂志 30 (2015), 501–513. DOI:<https://doi.org/10.1016/j.asoc.2015.01.048>伊万诺·德·法尔科·埃里克·拉斯科夫斯基、理查德·奥莱尼克、翁贝托·斯卡弗里、埃内斯托·塔伦蒂诺和马雷克·图德鲁伊。2015b.极值优化应用于分布式程序执行中的负载均衡。应用软计算 30 (2015), 501–513。

DOI:<https://doi.org/10.1016/j.asoc.2015.01.048>

奥克塔维奥·古铁雷斯·加西亚和阿德里安·拉米雷斯·纳法拉特。2015.云数据中心基于代理的负载均衡。簇

计算 18, 3 (2015), 1041–1062. DOI:<https://doi.org/10.1007/s10586-015-0460-x>

Saurabh Kumar Garg 和 Rajkumar Buyya。2011. Networkcloudsim:在云模拟中对并行应用程序进行建模。第四届 IEEE 公用事业和云计算国际会议 (UCC 11) 的会议记录。IEEE,105-113。

DOI:<https://doi.org/10.1109/UCC.2011.24>

Saurabh Kumar Garg, Steve Versteeg 和 Rajkumar Buyya。2013.云计算服务排名框架。

未来一代计算机系统 29, 4 (2013), 1012–1023. DOI:<https://doi.org/10.1016/j.future.2012.06.006>艾诺拉·贾法内贾德·戈

米·阿米尔·马苏德·拉赫马尼和努鲁丁·纳西·卡德尔。2017.云计算中的负载均衡算法:一项调查。网络与计算机应用杂志 88 (2017), 50–71。

DOI:<https://doi.org/10.1016/j.jnca.2017.04.007>

2017.04.007

穆罕默德·戈尼姆和拉利特·库尔卡尼。2017.用于可扩展异构系统的自适应 MapReduce 调度程序。

数据工程和通信技术国际会议论文集。施普林格,603-611。

DOI:https://doi.org/10.1007/978-981-10-1678-3_57

胡尔·戈什·弗朗西斯科·隆戈·维杰·K·奈克和基肖尔·S·特里维迪。2013.大规模 IaaS 云的建模和性能分析。未来一代计算机系统 29, 5 (2013), 1216–1234. DOI:<https://doi.org/10.1016/j.future.2012.06.005>

005

苏米·戈什和钱丹·巴纳吉。2016.云计算中基于优先级的改进节流算法。诉讼中

国际发明计算技术会议 (ICICT 16), 卷。3. IEEE,1-6。

希普拉·戈亚尔和马诺·库马尔·维尔马。2016.云计算环境中的负载均衡技术 综述。国际计算机科学与软件工程高级研究杂志 6, 4 (2016), 583–588。

桑迪普·古普塔、罗丝·罗宾·吉尔伯特、阿扬·班纳吉、扎赫拉·阿巴西·特里迪布·慕克吉和乔治·瓦萨莫普洛斯。2011年。

Gdcsim:用于分析绿色数据中心设计和资源管理技术的工具。国际绿色计算会议和研讨会 (IGCC 11) 的会议记录。IEEE,1-8。

哈坎·哈奇古姆斯、巴拉·艾耶和沙拉德·梅赫罗特拉。2002 年。提供数据库即服务。第 18 届 IEEE 国际数据工程会议论文集。IEEE,29-38。

DOI:<https://doi.org/10.1109/ICDE.2002.994695>韩一鸣和安东尼·西奥多·克罗诺普洛斯。2017.大规模集群和云系统的

可扩展循环自调度方案。国际并行编程杂志 45, 3 (2017), 595–611. DOI:<https://doi.org/10.1007/p10766-016-0434-5>

p10766-016-0434-5

迈克尔·霍尔·安东尼·西姆沃尼斯和大卫·伍德。2002.用于树负载均衡的维度交换算法。

西罗科 (2002), 181-196. DOI:<https://doi.org/10.1.1.18.2926>

Sue Chen Hsueh, Ming Yen Lin 和 Yi Chun Chiu。2014.一种负载均衡的 MapReduce 算法,用于具有多个键的基于块的实体解析。第 12 届澳大利亚并行和分布式计算研讨会论文集,第 152 卷。澳大利亚计算机学会,3-9。

Kai Hwang, Jack Dongarra 和 Geoffrey C. Fox。2013.分布式和云计算:从并行处理到物联网。摩根·考夫曼。

120:32

P. 库马尔和 R. 库马尔

- 伊戈尔·N·伊万尼森科 (Igor N. Ivanisenko) 和塔玛拉·A·拉德维洛娃 (Tamara A. Radivilova)。2015.分布式系统主要负载均衡算法综述。在
创新业务会议中的信息技术 (ITIB 15)。IEEE,89-92。
- 尼玛·贾法里,阿米尔·马苏德·拉赫马尼,艾哈迈德·哈比扎德·纳文和迈赫迪·侯赛因扎德。2015.专家云:基于云的共享人力资源知识和技能的框架。计算机在人类行为中的应用 46 (2015), 57-74。
- Yaser Jararweh,Zakarea Alshara,Moath Jarrah,Mazen Kharbutli 和 Mohammad N. Alsaleh。2013.Teachcloud:云计算教育工具包。国际云计算杂志 2, 2-3 (2013), 237-257。
- 布伦丹·詹宁斯和罗尔夫·施塔德勒。2015年。云中的资源管理:调查和研究挑战。网络与系统管理杂志 23, 3 (2015), 567-619。
- 马拉·卡尔拉和萨布吉特·辛格。2015.云计算中元启发式调度技术的回顾。埃及信息学杂志 16, 3 (2015), 275-295。
- 拉维·特贾·卡纳卡拉和维尤克·克里希纳·雷迪。2015.云计算环境下负载均衡技术的性能分析。IEEE 国际电气、计算机和通信技术会议 (ICECCT 15) 会议记录。IEEE,1-6。
- 姜秉锡和秋贤胜。2016.基于集群的大规模云去中心化作业调度。
EURASIP 无线通信和网络杂志 2016,25 (2016),1-8。
- Nidhi Jain Kansal 和 Inderveer Chana。2012。云负载平衡技术:迈向绿色计算的一步。国际
计算机科学杂志第 9 期, 1 (2012), 238-246。
- 玛扬卡·凯蒂尔和阿图尔·米什拉。2014.云计算环境下负载均衡算法的比较研究。
国际分布式和云计算杂志 1, 2 (2014), 5-14。
- 帕米特·考尔和希哈·梅塔。2017.使用增强型青蛙跳跃算法在云中进行资源配置和工作流调度。并行与分布式计算杂志 101 (2017), 41-50。
- 拉杰温德·考尔 (Rajwinder Kaur) 和帕万·卢特拉 (Pawan Luthra)。2012.云计算中的负载平衡。国际会议论文集
关于信息、电信和计算的最新趋势 (ITC 12)。Citeseer,374-381。
- 西娜·凯什瓦迪 (Sina Keshvadi) 和贝纳姆·法吉 (Behnam Faghhih)。2016. IaaS 云环境中基于多代理的负载均衡系统。国际机器人与自动化
杂志 1, 1 (2016), 1-6。
- 阿卜杜勒·纳西尔·汗、马特·基亚、萨米·汗和萨贾德·迈达尼。2013年。走向安全的移动云计算:一项调查。
未来一代计算机系统 29, 5 (2013), 1278-1299。
- 莱利·穆罕默德·汗利、赛亚德·纳赛尔·拉扎维和尼玛·贾法里·纳维米普尔。2008. LGR:用于网格计算系统的新的基于遗传的调度程序。IEEE 国际建模控制与自动化计算智能会议论文集。IEEE,639-644。
- Md. Tanzim Khorshed,ABM Shawkat Ali 和 Saleh A. Wasimi。2012年。关于云计算中的差距、威胁补救挑战以及主动攻击检测的一些想法的调查。未来一代计算机系统 28, 6 (2012),
833-851。
- Somayeh Kianpisheh,Nasrolah Moghadam Charkari 和 Mehdi Kargahi。2016. 基于蚁群的约束工作流程
异构计算系统的调度。集群计算 19, 3 (2016), 1053-1070。
- 芭芭拉·基钦纳姆。2004年。进行系统审查的程序。英国基尔,基尔大学 33,2004 (2004),1-26。
- 芭芭拉·基钦汉姆,O. Pearl Brereton,大卫·巴德根、马克·特纳、约翰·贝利和斯蒂芬·林克曼。2009年。软件工程的系统文献综述 系统的文献
综述。信息和软件技术 51, 1 (2009), 7-15。
- T. Kokilavani 和 DI George Amalarethinam。2011.静态元任务调度的负载平衡min-min算法
网格计算。国际计算机应用杂志 20, 2 (2011), 43-49。
- 埃图·库皮艾宁、米卡·曼蒂拉和尤哈·伊科宁。2015.在敏捷和精益软件开发中使用度量 系统的
工业研究的文献综述。信息和软件技术 62 (2015), 143-163。
- Jiayin Li, Meikang Qiu, Zhong Ming, Gang Quan, Xiao Qin, and Zonghua Gu。2012. Online optimization for scheduling
preemptable tasks on IaaS cloud systems. Journal of Parallel and Distributed Computing 72, 5 (2012), 666-677。
- Yu Liu, Changjie Zhang, Bo Li, and Jianwei Niu。2015. DeMS: A hybrid scheme of task scheduling and load balancing in
计算集群。网络与计算机应用杂志 83 (2015), 213-220。
- 弗拉维奥·隆巴迪和罗伯特·迪·彼得罗。2011。云计算的安全虚拟化。网络与计算机应用杂志 34, 4 (2011), 1113-1122。
- 拉坦·拉玛尼·马拉迪。2015.云计算中的负载平衡方法。国际创新研究杂志
科学、工程与技术 (在线) (2015), 2319-8753。
- 毛英迪、任道宁、陈曦。2013.云计算中基于预测模型的自适应负载均衡算法。第二届创新计算和云计算国际会议 (ICCC 13) 的会议记录。
ACM,纽约州,第 165 条,第 165-170 页。
- 丹·马里斯库、阿什坎·帕亚、约翰·莫里森和斯蒂芬·奥拉留。2017.一种扩展云资源管理的方法
ment。集群计算 20, 1 (2017), 909-924。

维克托·莫奇、马塞尔·昆泽和马吕斯·希伦布兰德。2013.高性能云计算。未来一代计算机系统 29, 6 (2013), 1408–1416。

卡洛斯·佩雷斯·米格尔、亚历山大·门迪布鲁和何塞·米格尔·阿隆索。2015年。对 Cassandra 的可用性进行建模。并行与分布式计算杂志 86 (2015), 29–44。

阿里雷扎·萨德吉·米拉尼 (Alireza Sadeghi Milani) 和尼玛·贾法里 (Nima Jafari)。2016.云环境中的负载均衡机制和技术:系统文献综述和未来趋势。网络与计算机应用杂志 71 (2016), 86–98。

Alireza Sadeghi Milani 和 Nima Jafari Navimipour。2016.云环境中的负载均衡机制和技术:系统文献综述和未来趋势。网络与计算机应用杂志 71 (2016), 86–98。

纳德·穆罕默德、贾米拉·阿尔·贾鲁迪和阿卜杜拉·艾德。2013年。一种用于快速文件下载的双向技术,具有云中的动态负载均衡。网络与计算机应用杂志 36, 4 (2013), 1116–1130。

鲁帕姆·穆霍帕迪亚 (Rupam Mukhopadhyay)、迪比亚约蒂·戈什 (Dibyajyoti Ghosh) 和南迪尼·穆克吉 (Nandini Mukherjee)。2010.现有负载均衡算法在大型、动态、异构分布式系统中的应用研究。第九届国际软件工程、并行和分布式系统会议论文集。世界科学工程学院和学会 (WSEAS), 238–243。

拉内什·库马尔·纳哈和穆罕默德·奥斯曼。2016年。云中成本感知服务代理和性能感知负载均衡算法。网络与计算机应用杂志 75 (2016), 47–57。DOI:<https://doi.org/10.1016/j.jnca.2016.08.018>

尼玛·贾法里·纳维米普尔。2015.规范和验证值得信赖的人力资源的方法

专家云中的发现机制。专家系统与应用 42, 15 (2015), 6112–6131。

尼玛·贾法里·纳维米普尔和耶加内·查班德。2016.项目团队的知识共享机制和技术:文献综述、分类和当前趋势。计算机在人类行为中的应用 62 (2016), 730–742。DOI:<https://doi.org/10.1016/j.chb.2016.05.003>

尼玛·贾法里·纳维米普尔和法纳兹·谢里菲·米拉尼。2015.资源发现技术综合研究

点对点网络。点对点网络和应用程序 8, 3 (2015), 474–492。

乔瓦尼·内利亚、马泰奥·塞雷诺和朱塞佩·比安奇。2016.跨绿色数据中心的地理负载均衡:A

平均场分析。ACM SIGMETRICS 性能评估评论 44, 2 (2016), 64–69。

Kumar Nishant, Pratik Sharma, Vishal Krishna, Chhavi Gupta, Kuwar Pratap Singh, Ravi Rastogi 等。2012.使用蚁群优化对云中节点进行负载均衡。第 14 届国际计算机建模与仿真会议 (UKSim) 论文集。IEEE, 3–8。

克莱塞姆·阿尔·努埃米、纳德·穆罕默德、玛丽亚姆·阿尔·努埃米和贾米拉·阿尔·贾鲁迪。2012年。云计算中的负载均衡调查:挑战和方法。在 IEEE 第二届网络云计算和应用研讨会 (NCCA 12) 中。IEEE, 137–142。

阿尔贝托·努涅斯、何塞·路易斯·巴斯克斯·波莱蒂、奥古斯丁·C·卡米内罗、耶稣·卡雷特罗和伊格纳西奥·马丁·略伦特。2011.新型云计算仿真平台设计。计算科学及其应用国际会议论文集。施普林格, 582–593。

西蒙·奥斯特曼、卡西安·普兰肯斯坦纳、拉杜·普罗丹和托马斯·法林格。2010. GroudSim:用于计算网格和云的基于事件的模拟框架。欧洲并行处理会议论文集。

施普林格, 305–313。

埃琳娜·帕西尼、克里斯蒂安·马特奥斯和卡洛斯·加西亚。2015.通过蚁群优化平衡在线科学云的吞吐量和响应时间。工程软件进展 84 (2015), 31–47。DOI:<https://doi.org/10.1016/j.advengsoft.2015.01.005>

努斯拉特·帕夏·阿米特·阿加瓦尔和拉维·拉斯托吉。2014.云计算环境中虚拟机负载均衡算法的循环方法。国际计算机科学和软件工程高级研究杂志 4, 5 (2014), 34–39。

兹吉斯瓦夫·帕夫拉克。2001.粗糙集和决策算法。计算机科学讲义卷。2005 (2001), 30–45。

DOI:https://doi.org/10.1007/3-540-45554-X_3

Florin Pop, Ciprian Dobre, Valentin Cristea, Nik Bessis, Fatos Xhafa 和 Leonard Barolli。2015.云环境中非周期性任务的截止日期调度:一种新的资源管理方法。超级计算杂志 71, 5 (2015), 1754–1765。

鲁帕利·旁杰和拉克什·库马尔。2018.健康监测 WBAN 的技术方面:全面审查。

无线网络 (2018), 1–33。

法希梅·拉梅扎尼、卢杰和法鲁克·哈迪尔·侯赛因。2014.使用粒子群优化的云计算中基于任务的系统负载均衡。国际并行编程杂志 42, 5 (2014), 739–754。DOI:<https://doi.org/10.1007/s10766-013-0275-4>

马丁·兰德斯、大卫·兰姆和 A.塔勒布·本迪亚布。2010.云计算分布式负载均衡算法的比较研究。第 24 届 IEEE 国际高级信息网络和应用研讨会会议记录, WAINA。IEEE, 551–556。

苏米亚·雷和阿旂塔·德·萨卡。2012.云计算环境下负载均衡算法的执行分析。

国际云计算杂志:服务和架构 2, 5 (2012), 1-13。

玛丽亚·A·罗德里格斯和拉吉库马尔·布亚。2017.在多租户科学工作流程中调度动态工作负载作为服务平台。未来一代计算机系统 79, 2 (2017), 739-750。

Mohsen Amini Salehi, Jay Smith, Anthony A. Maciejewski, Howard Jay Siegel, Edwin KP Chong, Jonathan Apodaca, Luis D. Briceno, Timothy Renner, Vladimir Shestak, Joshua Ladd 等。2016.异构计算系统中独立任务的基于随机的鲁棒动态资源分配。并行与分布式计算杂志 97 (2016), 96-111。

沉海英,于雷,陈六华,李卓兆。2016年.告别固定带宽预留:云中具有弹性带宽预留的作业调度。IEEE 国际云计算技术与科学会议 (CloudCom 16) 论文集。IEEE,1-8。

施建辉,孟春雷,马玲丽。2011.基于混合调度的分布式负载均衡策略。第四届计算科学与优化国际联合会议 (CSO 11)。IEEE,268-271。

阿尔蒂·辛格·迪普尔·朱尼贾和马尼沙·马尔霍特拉。2015.云计算中基于自治代理的负载均衡算法。Procedia 计算机科学 45 (2015), 832-841。

朴雅卡·辛格·帕拉克·巴加和索拉布·古普塔。2016.云计算中的各种负载平衡算法:一项调查。

国际计算机应用杂志 143, 7 (2016), 1-8。

普纳姆·辛格·梅特里耶·杜塔和纳文·阿加瓦尔。2017.云计算中基于元启发式方法的调度任务综述。知识与信息系统 (2017) ,1-51。

苏赫帕尔·辛格和因德维尔·查纳。2015.云计算中的 QoS 感知自主资源管理:系统化审查。计算机调查 48, 3,第 42 条 (2015),42 页。DOI:<https://doi.org/10.1145/2843889>

艾丽卡·索萨·费尔南多·林斯·爱德华多·塔瓦雷斯·保罗·库尼亚和保罗·马西尔。2015.考虑可靠性和成本要求的云基础设施规划建模方法。

IEEE 系统、人类和控制论汇刊:系统 45, 4 (2015), 549-558。

苏巴希尼·苏巴希尼和维拉鲁纳·卡维萨。2011.云计算服务交付模型安全问题调查。网络与计算机应用杂志 34, 1 (2011), 1-11。

M. Suresh.S. Karthik 和 B. Santosh Kumar。2014.使用ANFIS和GSO的公共云负载均衡模型。在智能计算应用国际会议 (ICICA 14) 的会议记录中。IEEE ,85-89。

L·塔斯奎尔。2015.适用于多云环境的基于代理的负载均衡器。哥伦比亚国际出版杂志云计算研究 1, 1 (2015), 35-49。

技术。2017.2017年十大云计算服务提供商。2018年2月20日检索自<https://www.technavio.com/blog/top-10-cloud-computing-service-providers-2017>。

威廉·沃斯鲁伊斯·詹姆斯·布罗伯格和拉吉库马尔·布亚。2011.云计算简介。云计算:原理和范式 87 (2011), 1-44。

王伟仁、张岳山、罗温宗和李毅康。2013.具有QoS的并行任务自适应调度

对混合云环境的满意度。超级计算杂志 66, 2 (2013), 783-811。

王振华、陈浩鹏、付英、刘德林、班云梦。2015.云上快速存储系统的工作负载平衡和自适应资源管理。未来一代计算机系统 51 (2015), 120-131。

Bhathiya Wickremasinghe, Rodrigo N. Calheiros 和 Rajkumar Buyya。2010. Cloudanalyst:基于 Cloudsim 的可视化建模器,用于分析云计算环境和应用程序。第 24 届 IEEE 高级信息网络和应用国际会议 (AINA) 论文集。IEEE,446-452。

Tin Yu Wu, Wei Tsong Lee, Yu San Lin, Yih Sin Lin, Hung Lin Chan 和 Jihh Siang Huang。2012.基于云存储的动态负载均衡机制。IEEE 国际计算、通信和应用会议论文集。IEEE,102-106。

吴志龙,邢盛,蔡树斌,肖志姣,钟明。2016.云系统任务调度的遗传-蚁群混合算法。智能计算与通信国际会议论文集。

施普林格,183-193。

肖震、宋伟佳、陈琪。2013.云计算环境中使用虚拟机的动态资源分配。IEEE 并行和分布式系统汇刊 24, 6 (2013), 1107-1117。DOI:<https://doi.org/10.1109/TPDS.2012.283>

于欣,谢志强,杨静。2017.一种面向负载均衡的并行任务成本高效调度方法。

网络与计算机应用杂志 81 (2017) ,37-46。

Moona Yakhchi, Seyed Mohssen Ghafari, Shahpar Yakhchi, Mahdi Fazeli 和 Ahmad Patooghi。2015.提出一种基于布谷鸟优化算法的负载均衡方法,用于云计算基础设施中的能源管理。

第六届建模、仿真和应用优化国际会议 (ICMSAO 15) 论文集。IEEE,1-5。

杨吉祥、凌凌、刘海滨。2016.考虑大型分布式计算系统通信延迟开销的分层负载均衡策略。工程数学问题 2016 (2016), 1-9。

Faouzia Zegrari,Abdellah Idrissi 和 Hajar Rehioui. 2016.云环境中高效负载均衡的资源分配。大数据和先进无线技术国际会议 (BDAW 16) 论文集。ACM,纽约州,第 46 条,7 页。 DOI:<https://doi.org/10.1145/3010089.3010131>

Lingfang Zeng,Bharadwaj Veeravalli 和 Albert Y. Zomaya. 2015.云环境下工作流应用的集成任务计算和数据管理调度策略。网络与计算机应用杂志 50 (2015), 39–48。

张奇,程鲁,拉乌夫·布塔巴。 2010。云计算:最先进的技术和研究挑战。杂志 互联网服务和应用 1, 1 (2010), 7–18。

赵佳,杨坤,魏晓晖,丁彦,胡亮,徐高超。 2016.一种基于启发式集群的云环境中使用贝叶斯定理负载均衡的任务部署方法。 IEEE 并行和分布式系统汇刊 27, 2 (2016), 305–316。

迪米特里奥斯·齐西斯和迪米特里奥斯·莱卡斯。 2012年。解决云计算安全问题。下一代计算机系统 28, 3 (2012), 583–592。

2017年12月收到； 2018年8月修订； 2018年8月接受