



计算机科学

COMPUTER SCIENCE

多云 workflow 调度综述

于浩雯, 刘波, 周娜琴, 林伟伟, 柳鹏

引用本文

于浩雯, 刘波, 周娜琴, 林伟伟, 柳鹏. [多云 workflow 调度综述](#)[J]. 计算机科学, 2022, 49(11): 250-258.

YU Hao-wen, LIU Bo, ZHOU Na-qin, LIN Wei-wei, LIU Peng. [Survey of Multi-cloud Workflow Scheduling](#)[J].

Computer Science, 2022, 49(11): 250-258.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[混合云 workflow 调度综述](#)

Survey of Hybrid Cloud Workflow Scheduling

计算机科学, 2022, 49(5): 235-243. <https://doi.org/10.11896/jsjcx.210300303>

[边缘环境下基于模糊理论的科学 workflow 调度研究](#)

Study on Scientific Workflow Scheduling Based on Fuzzy Theory Under Edge Environment

计算机科学, 2022, 49(2): 312-320. <https://doi.org/10.11896/jsjcx.201000102>

[高速公路中的移动边缘计算研究](#)

Research on Mobile Edge Computing in Expressway

计算机科学, 2021, 48(6A): 383-386. <https://doi.org/10.11896/jsjcx.200900212>

[一种应急指挥系统的设计与实现](#)

Design and Implementation of Emergency Command System

计算机科学, 2021, 48(6A): 613-618. <https://doi.org/10.11896/jsjcx.201000136>

[基于 jBPM 的科学试验管理系统的设计与实现](#)

Design and Implementation of Scientific Experiment Management System Based on jBPM

计算机科学, 2021, 48(6A): 658-663. <https://doi.org/10.11896/jsjcx.200600158>

多云 workflow 调度综述

于浩雯¹ 刘波¹ 周娜琴² 林伟伟³ 柳鹏¹

¹ 华南师范大学计算机学院 广州 510631

² 广州大学网络空间先进技术研究院 广州 510006

³ 华南理工大学计算机科学与工程学院 广州 510006

(me@yuhaowen.com)

摘要 传统的云供应商单独为用户提供服务,这导致了本地云资源不足和扩展费用较高等问题。而新兴的多云组合地理位置不同的云供应商的服务,为用户提供了更多的选择,逐渐成为了研究的热点。同时,workflow 调度又是多云研究的关键问题之一。为此,文中首先对多云环境下的 workflow 调度技术做了深入的调查和分析,然后将多云下的 workflow 调度方法进行分类和比较,重点阐述了面向成本、面向完工时间的单目标优化 workflow 调度,面向成本和完工时间,面向响应时间和成本,面向可靠性、成本和完工时间的多目标优化 workflow 调度,以及面向其他多目标优化的多云 workflow 调度。最后,在此基础上讨论了多云环境下 workflow 调度的未来研究方向:不确定性 workflow 调度、能耗与其他目标的联合调度优化、与边缘服务器协同的调度优化、虚拟机和 Serverless 平台混合调度。

关键词: 多云; workflow; 资源调度

中图法分类号 TP393

Survey of Multi-cloud Workflow Scheduling

YU Hao-wen¹, LIU Bo¹, ZHOU Na-qin², LIN Wei-wei³ and LIU Peng¹

¹ School of Computer Science, South China Normal University, Guangzhou 510631, China

² School of Computer Science and Network Engineering, Guangzhou University, Guangzhou 510006, China

³ School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

Abstract Traditional cloud providers provide services solely for users, which has problems of insufficient local cloud resources and high expansion costs. While the emerging multi-cloud combines services of cloud providers with different geographical locations providing users with more choices and has gradually become a research hotspot. At the same time, workflow scheduling is one of the key problems in multi-cloud research. Therefore, this paper firstly makes a thorough investigation and analysis on workflow scheduling technology in multi-cloud environment, then compares and classifies the workflow scheduling method included. This paper focuses on single objective workflow scheduling optimization problem for cost and completion time, multi-objective optimization workflow scheduling problem for cost and completion time, for response time and cost, for reliability, cost and completion time, and for other multi-objective optimization. Finally, the future research directions of workflow scheduling in the multi-cloud environment are discussed, including uncertain workflow scheduling, joint scheduling optimization of energy consumption and other objectives, scheduling optimization with edge servers, and hybrid scheduling of virtual machine and Serverless platform.

Keywords Multi-cloud, Workflow, Scheduling of resources

1 引言

使用模式为各种组织和客户提供各种 IT 服务^[1]。它通过虚拟化技术对外提供了一个几乎无边无际的存储资源池,为人们存储和访问资源提供了便利。尽管云中的资源对用户来说

云计算是一种基于网络的不断发展的技术,旨在根据付费

到稿日期:2021-12-21 返修日期:2022-05-20

基金项目:广东省重点领域研发计划(2021B0101420002,20200704002);国家自然科学基金面上项目(62002078,61872084);广东省基础与应用基础研究基金(2019B030302002);广州市开发区国际合作项目(2020GH10,2021GH10)

This work was supported by the Key Research and Development Program of Guangdong Province(2021B0101420002,20200704002), General Project of National Natural Science Foundation of China(62002078,61872084), Basic and Applied Basic Research Foundation of Guangdong Province(2019B030302002) and Guangzhou Development Zone International Cooperation Project(2020GH10,2021GH10).

通信作者:刘波(liugubin@sohu.com)

是有限的,但是单个云供应商能够给用户提供的资源是有限的,仅停留在使用单个供应商提供的服务存在着服务不可用、供应商锁定、隐私泄露等风险^[2]。一些包括服务器停机、自然灾害、电源故障等的不可预知事件可能会导致服务不可用,使得依赖单个云供应商服务的用户蒙受巨大的损失。同时,锁定单个供应商使得用户容易受到价格上涨、可用性下降,甚至供应商破产的影响。而放弃原供应商,又不得不考虑迁移至新供应商的巨大成本。再次,将用户数据完全暴露在单个云供应商中,一旦单个云供应商受到外部攻击,就会加大用户隐私泄露带来的负面影响。

多云的出现为降低这些风险提供了很好的解决方案,它正成为一种新兴范式^[3]。根据 IBM 的权威定义,多云指使用来自不同云提供商的两个或多个云,可以是基础设施、平台或软件即服务(IaaS, PaaS 或 SaaS)的任意组合^[4]。也就是说,多云通过使用多个云而不是一个云,将需要水平伸缩的应用程序和服务拓展到具有不同定价模型的云上。从多云的权威定义可以看出,与单云相比,多云组合了地理位置不同的云供应商的服务,因此降低了用户对单个云的依赖程度,一旦单个云发生服务中断,多云中的其他云会继续为用户提供服务^[5]。同时,用户可以选择性价比较高的云服务组合^[6]。再者,用户的数据分散存储在不同云中,这降低了单个云隐私泄露带来的负面影响。因此,在单个云供应商提供资源有限的情况下,合理选择多云提供的组合服务是较为明智的选择^[7]。GLOBE NEWSWIRE 的云计算研究报告表明,2019 年全球云计算市场价值达到 3 210 亿美元,预计到 2026 年将达到 10 259 亿美元^[8],其中多云也将成为主流。

在云计算领域,对于大规模优化问题,科学工作流调度是 NP-hard 问题^[9],这是科学工作流中相互依赖任务的复杂性决定的。针对这一问题, Tomarchio 等^[10]系统地回顾了多云环境下的软件框架,并从云和应用程序的角度分析每一类框架的主要功能,结果表明多云环境中更需要跨多种云部署考虑自组织、自管理和自修复的或具有连续交付功能的自主管理系统。Hosseinzadeh 等^[11]以元启发式多目标优化为背景,根据应用的算法对各种云环境下的多目标调度方法进行了

分类综述。现有研究资源调度的相关文献^[12-14]并不完全针对多云环境,而是往往与单云、混合云等合并阐述,其研究的调度单位往往没有细化为工作流,而是与独立任务调度混杂。鉴于目前还没有针对多云工作流调度的综述研究,本文对多云环境下的工作流调度技术做了深入的调查和分析,重点阐述了多云环境下面向成本、面向完工时间的单目标优化工作流调度以及面向多目标优化的工作流调度,进而讨论了未来研究的方向和趋势。

2 相关背景

2.1 多云介绍

多云环境由若干不同 IaaS 云服务器供应商组成,这些供应商可提供不同价格和性能的虚拟机,但彼此独立、互不通信^[15]。而不同的供应商采用的计费机制不同,例如云提供商 Amazon EC2 采用小时为单位的粗粒度计费周期,不足整小时算作整小时计费^[16],云服务提供商微软 Azure 则采用分钟为单位的计费周期^[17]。这意味着用户利用多云合理的组合计费机制,可以选择最适合自己的计费粒度,从而节约成本。

多云环境下,每个云平台都装有独立的资源管理器,向全局调度器实时发送当前虚拟机的资源情况。而全局调度器通过外部代理或统一的 API 接口获得外部云平台发送过来的资源信息,根据当前任务情况将任务分段传输给不同云进行处理^[14]。这些任务通常安排在多个 IaaS 云提供商上,以协作方式完成应用程序。这种方式增加了任务处理的并行性,提高了工作效率。图 1 给出了一个多云环境的全局视图,每个云都独立于其他云,并且以不同的成本提供不同级别的计算能力。相同云中计算资源之间的网络带宽基本不变,而不同云之间的网络带宽是不同的、可变的。另外,云中的计算资源之间的延迟通常很低,而不同云之间的延迟则相对较高^[18]。因此,通常的做法是,尽量将依赖性较强的任务部署到相同云中,降低带宽不稳定和延迟带来的负面影响,同时将依赖性弱的任务部署到不同云中,增加任务执行的并行性,以保持较高的工作效率。

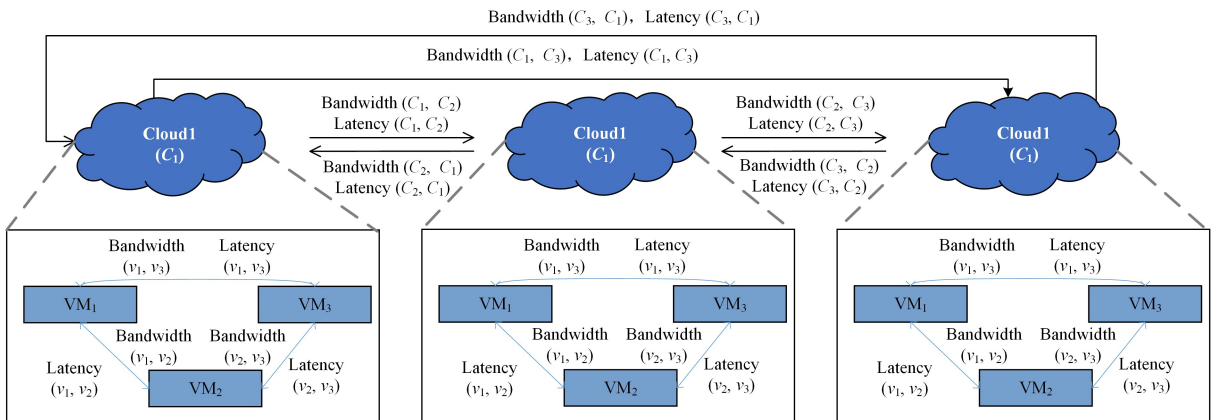


图 1 多云环境的全局视图

Fig. 1 Global view of multi-cloud environment

对于用户而言,多云为用户提供了更加灵活的服务。云的多种形式使得各种基础设施得到了扩展,用户可以根据自己的需求选择合适的组合服务。对于云供应商而言,多云

可以将原本需要在单个云上完成的工作交由多个云共同完成。这种方式降低了单个云的工作压力,同时增加了处理工作的并行性,使得任务的完成更为高效。

2.2 workflow 调度

随着云技术的发展和云平台的推广,云计算领域的工作流调度问题成为了一个重要的研究课题^[19]。工作流调度包括两个不可分割的部分,即资源发放和任务调度^[20]。前者指资源注册的决策,包括类型、数量和使用时间;后者是将所有任务映射到已登记的资源,并对每个资源上的任务进行排序。考虑到工作流中任务之间的依赖关系,在资源映射时必须先处理父任务,再处理子任务,同时还要贴合用户的需求,因此云中的工作流调度通常为 NP-hard 问题^[21]。一个工作流可能包含数百个甚至更多的任务,云平台上的可行解决方案随着任务数量的增加呈指数增长,搜索最优调度的成本通常非常高。因此,设计具有低复杂度的启发式算法或元启发式算法成为了在合理时间内生成高效调度的一个有吸引力的选择^[22]。

工作流调度技术与能够提供强大计算能力的云结合,被广泛应用于各类工作流。云计算领域的工作流由一系列相互依赖的任务构成,通常由有向无环图(Directed Acyclic Graph, DAG)表示^[23]。云计算中的工作流通常划分为动态工作流和静态工作流^[24]。其中,流式工作流是动态工作流的代表,而静态工作流有业务工作流(Business Workflow)和科学工作流(Scientific Workflow)之分。业务工作流通常用来表示商业应用程序,而科学工作流用于天文学、生物信息学、地震科学和物理学等领域。典型的科学工作流包括:天文学领域的 Montage、重力物理学领域的 LIGO、地震科学领域的 CyberShake、生物信息学领域的 SIPHT 和生物基因学领域的 Epigenomics 等^[25]。为了更加形象地描绘这些经典科学工作流,图 2 给出了这些工作流的形态。这些工作流具有不同的特点,如流式工作流对实时性要求较高,需要就近分配虚拟机资源。科学工作流通常是数据密集型的,需要云提供商为客户提供更多的虚拟机类型^[26],因此单一的云服务很难满足他们的需求。多云范例则可以有效地解决这一问题,通过组合服务的方式为不同工作流提供更丰富的虚拟机资源,更贴合用户需求的服务。

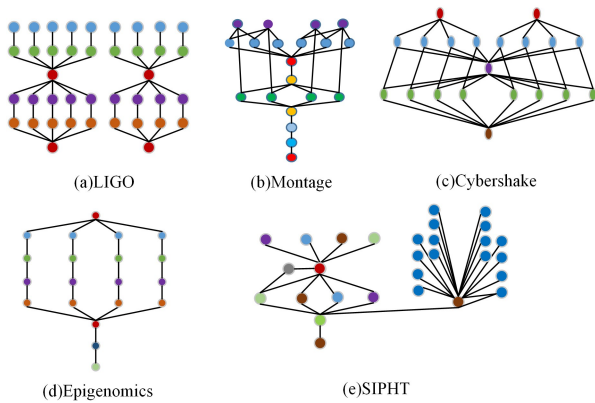


图 2 来自不同研究领域的 5 个著名科学工作流的例子^[25]

Fig. 2 Five examples of well-known scientific workflows from different areas of research^[25]

2.3 QoS 调度约束

工作流调度问题十分复杂,目前很难实现最优调度^[27]。因此,通常情况下,人们只能在 QoS 约束下为工作流选择适当的虚拟机资源^[28],即在满足用户要求的同时,选择一个或

一组指标进行工作流调度算法的优化,然后通过服务水平协议(Service Level Agreement, SLA)对调度质量进行评估。本文对近五年多云工作流调度的文献进行整理,将 QoS 约束指标按出现频次由高到低进行排列,即成本、完工时间、响应时间、可靠性、能耗、吞吐量、资源利用率等,如表 1 所列。

表 1 工作流调度的 QoS 约束

Table 1 QoS constraints for workflow scheduling		
出现占比	QoS 约束	描述
44%	成本	用户为云供应商提供的服务所支付的总费用
26%	完工时间	从工作流调度开始到该工作流完成执行所需的时间
10%	响应时间	用户发出应用程序请求到他们收到相应响应的的时间
8%	可靠性	描述了供应商能够满足用户需求的概率
5%	能耗	决定了调度过程中能量的利用率
5%	吞吐量	供应商在一定时间内满足用户请求的能力
2%	资源利用率	资源被有效利用的时长和总占用时长的比值

3 多云的工作流调度

众所周知,云工作流调度是 NP-hard 问题^[29],而多云环境下的工作流调度则更加复杂^[30]。这是因为,多云环境中的服务是由多个独立的云平台提供的,从多个云平台中选择最优服务组合来满足 QoS 要求是一项具有挑战性的工作。本节结合多云工作流调度中高频出现的 QoS 约束指标,首先将近几年多云下工作流的调度方法划分为面向单目标的工作流调度和面向多目标的工作流调度,然后根据具体目标对文献进行分类综述,最后将涉及的调度方法按照优化目标、方法描述、关键实现、工具和特点 5 个维度进行综合分析和对比。

3.1 面向单目标优化的多云工作流调度

单目标工作流调度(Single-objective Workflow Scheduling)指在工作流调度过程中只考察单个目标,以对整个工作流调度进行优化。由于仅仅考察单个目标,优化效果通常较为明显,但在现实场景中往往需要同时对多个目标进行优化,因此针对单目标工作流调度的研究较少。本节将面向单目标的工作流调度分为面向成本和面向完工时间的多云工作流调度。

3.1.1 面向成本的多云工作流调度

传统的集群计算框架在异构云上的可扩展性和适应性有限,已难以应对前所未有的数据量庞大、数据的地理分散分布、数据跨云分布的状况。在多云环境下,随意跨云运行数据密集型应用程序的效率非常低,可能会产生惊人的开销。云支出主要包括出口网流量费用和订阅费用,而云支出过多主要是由出口数据传输过多和主机资源利用率不足引起的^[31]。为了解决出口数据传输过多的问题,Jiang 等^[31]开发了一种新颖的成本感知调度算法。该算法采用了数据局部性推理方法,通过将计算资源放置在数据附近,即将任务放置在具有其输入数据的主机上或放在与该主机同云的主机上,来减少数据移动占用的带宽,降低出口网络流量成本。但该算法未设置任务抢占式虚拟机的规则,因此不能从降低虚拟机订阅成本的角度降低总成本。

为了从提升主机利用率的角度降低成本,Jiang 等^[32]将任务按照依赖关系分组后按资源需求量递减排序,将来自多个云的主机按照容量大、带宽充裕、出入口成本低优先的原则排序,然后按照 First-Fit(FF)原则将分组的任务分配给主机。该算法使主机资源利用率得到提高,从而显著降低了虚拟机的租赁成本。

3.1.2 面向完工时间的多云工作流调度

在面向完工时间的多云工作流调度研究中,当前的方法主要是先将工作流按照任务之间的依赖程度划分为块,再根据每个任务块的预测执行时间提出启发式算法或元启发式算法来调度工作流,使任务块在多个云上并行执行,从而缩短工作流调度的完工时间。Xu 等^[33]为了保证调度方案在期限内完成,提出了一种新颖的调度算法。它首先根据科学工作流自身的结构特点,尽可能地合并存在有向割边的任务,从而减少总的任务数,这相当于缩短了总时间;然后不断把每个关键父任务作为下一次输入的子任务继续寻找关键父任务,直至所有路径覆盖全部任务,任务根据依赖关系被排列好后,依次执行;最后把每个局部关键路径分配到最适合的云上并行执行任务,以获得每一个任务的实际结束时间。该算法通过合并任务和科学地排列任务顺序,减小了虚拟机的处理压力,缩短了完工时间。

合理规划工作流中任务的执行顺序固然能增加任务在多云环境下执行的并行性,而合理利用工作流执行的中间数据则能够进一步提高工作流的执行效率。Heidsieck 等^[34]提出了 3 种缓存感知的分布式调度算法,即 CacheA, Sgreedy-CH 和 Fgreedy-CH,它们试图通过缓存工作流中间数据或让工作流在多个云中并行执行的方式来缩短工作流调度的完工时间。CacheA 是一种新的贪婪算法,它执行缓存感知调度,另外两种算法将分布式贪婪调度算法^[35]扩展为缓存感知算法,这 3 种算法通过有选择缓存和重用中间数据的方式进行自适应缓存、缓存站点选择和动态工作流调度。Heidsieck 等提出的方法虽然牺牲了额外空间存储工作流的缓存,同时还需要维护缓存索引,但总体而言,仍然提高了工作流的执行效率,从而缩短了工作流调度的完工时间。同时,在应对工作负载难以估计的工作流或计算能力在执行期间变化很大的环境下,该算法在缩短完工时间方面更有优势。

3.2 面向多目标优化的多云工作流调度

多目标工作流调度(Multi-objective Workflow Scheduling)指在工作流调度过程中同时考察多个目标,这几个目标在通常情况下是相互冲突的,也就是说,不存在同时优化所有目标的解决方案,只能根据需求设定不同目标对最终结果的影响程度,让调度方案总体在选定目标上维持较高的水平。由于面向多目标工作流调度比面向单目标工作流调度更加符合现实场景,该分类下的文献比单目标下的文献更加丰富。本节将面向多目标的多云工作流调度分为以下几类:面向成本和完工时间,面向响应时间和成本,面向可靠性、成本和完工时间,面向其他多目标优化的多云工作流调度。

3.2.1 面向成本和完工时间的多云工作流调度

成本和完工时间是多云工作流调度的重要指标,因此多云工作流调度中大部分工作仅将完工时间和成本作为调度目标,

采用的方法也多种多样。与 Xu 等^[33]提出的方法相似,为了缩短云间传输时间,Senturk 等^[36]也将 DAG 关键路径中依赖性较强的任务分为同组。与 Xu 等提出的方法不同,Senturk 等提出的方法不是将任务组在虚拟机上的执行时间短作为任务组的优先分配条件,而是将执行时间和成本的比值纳入虚拟机的选择条件,从而得到成本和完工时间较优的分配方案。

与 Jiang 等^[31]提出的方案一样,Ulabedin 等^[37]看到了利用任务和数据的依赖关系能够减少数据传输,以及减少数据传输对降低云间传输成本的作用。但 Ulabedin 等不是通过将任务组放置在与输入数据同云的主机上来减少数据移动,而是针对跨数据中心却需要相同数据集的任务组,采用任务复制的方式来提高数据可用性,从而减少数据移动。如图 3 所示,为了利用固定数据集 d_4 ,将任务 T_6 从数据中心 DC-2 复制到 DC-3,从而减少了将数据集 d_4 移动到 DC-3 的数据传输成本。由于数据集通常体量较大,因此移动数据集需要大量的时间。Ulabedin 等不仅降低了出口网络流量成本,也缩短了由于数据集复制而花费的大量时间,从而提高了工作流在多云环境下的执行效率。

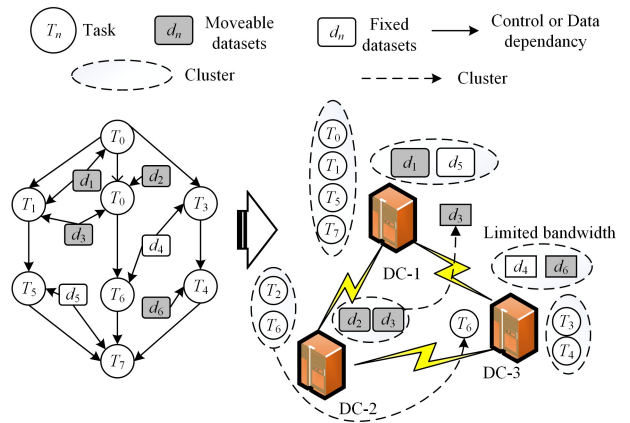


图 3 跨数据中心的任务间通信^[37]

Fig. 3 Intertask communication across data centers^[37]

在面向成本和完工时间的多目标优化中,Senturk 等和 Ulabedin 等提到的方法固然是优秀的方法,但没有考虑所得解的稳定性,即成本和完工时间在工作流调度中的比重被更改之后,之前建立的求解模型是否能保持良好表现的问题。为了研究所得解的稳定性,Mohammadi 等^[27]从数学角度解决了多云下的多目标优化问题。该方法通过对完工时间和成本这两个单目标进行线性组合,来构成新的目标函数,然后通过更改目标函数中两个单目标的权值,来得到一组可行解,最后对目标函数不同权重下产生的结果均值进行后优化分析,最终得到可行解中最稳定的解。该方案虽然在解的稳定性上表现良好,且一定程度地减小了成本和缩短了完工时间,但线性函数中的权值需要具有先验知识的人参与决定,这增加了该方案的人工成本,而文献^[27]并未将这部分成本加入到最终的评价体系。

3.2.2 面向响应时间和成本的多云工作流调度

在多云中,用户/数据和不同位置的服务之间的网络延迟可能会显著影响应用程序的性能^[38]。为了创建更高质量的应用程序,研究人员不应局限于缩短调度和执行工作流的完工时间,而且开始关注响应时间,即从用户发出应用程序请求

到他们收到相应响应的的时间。目前已有方法通常的思路为,将应用程序部署到更接近大多数用户的位置,从而缩短总体响应时间,但这样会增加部署虚拟机的数量,从而增加成本。为了实现响应时间和成本的折中调度,现有的研究通常采用染色体编码部署的方案,通过多目标进化算法从多个云供应商中选择合适的资源,从而提高解的质量。Ma 等^[39]提出了使用 NSGA-II 和局部搜索结合的方法,并采用两个与响应时间和成本相关的函数来约束种群的进化方向。但 NSGA-II 算法不可能在大规模优化问题中寻找真正的帕累托最优,只能找到尽可能接近帕累托最优的解决方案。因此,为了寻求更优解,Ma 等将 NSGA-II 算法与局部搜索相结合,在每一代

中发现更接近帕累托前沿的解。该方法在使用相同计算资源的情况下,提升了算法的性能。

Ma 等对每一代的后代种群进行局部搜索,从而进一步提高进化过程中的后代质量,但忽视了提高初始化种群的质量。而多目标进化算法的性能很大程度上依赖于种群初始生成初始种群的质量^[40]。Shi 等^[41]提出的基于种子的多目标进化算法很好地解决了该问题。如图 4 所示,Shi 等将该问题视作两个单目标问题,生成两个响应时间最优和成本最优的种子,然后通过均匀交叉等方式拓展初始种群,构建了更加广泛的高质量初始种群。高质量的种群使得 BB-Seed 生成了显著缩短平均响应时间和降低成本的调度方案。

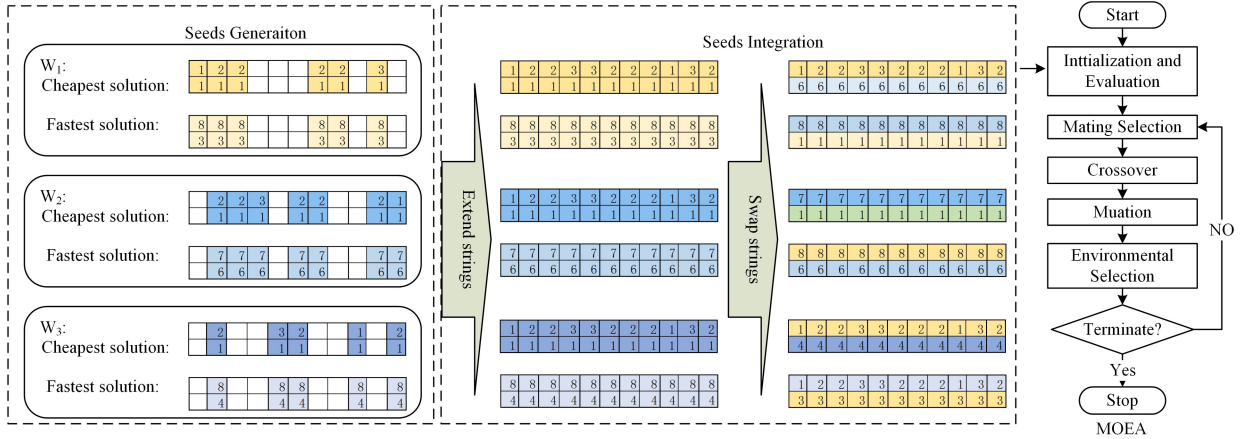


图 4 基于种子的多目标进化算法^[41]

Fig. 4 Seeding-based multi-objective evolutionary algorithms^[41]

3.2.3 面向可靠性、成本和完工时间的多云 workflow 调度

与其他分布式系统一样,云计算也容易出现软件故障、硬件故障和电源故障。在执行复杂的工作流应用程序过程中,这些不可避免的问题会导致工作流执行失败^[42]。因此,多云环境下调度工作流在优化其他目标的同时确保可靠性是非常重要的^[43]。到目前为止,多云 workflow 调度中仅将完工时间和成本作为调度目标的居多。为了在最小化完工时间和成本的同时将可靠性也纳入多云 workflow 调度优化中,Hu 等^[44]采用了粒子群优化(Particle Swarm Optimization, PSO)算法。他们把工作流中依赖性较强的任务视作粒子,将不同云供应商的可靠性、成本和完工时间同时纳入粒子的移动参考。为了约束粒子的活动范围,设计了限制粒子越界的函数,通过越界函数将可靠性、成本和完工时间超出限制的粒子移除,从而得到最终的调度方案。

虽然粒子群算法具有容易实现、收敛较快的优点,但同时存在容易陷入局部最优解的缺点。Farid 等^[45]则针对这一问题改进了 PSO 算法,提出了一种基于模糊资源利用的多目标调度算法(Multi-objective Scheduling Algorithm With Fuzzy Resource Utilization, FR-MOS)。该算法不仅将成本和完工时间纳入粒子的移动参考,而且独创性地利用模糊系统^[46]与资源利用率来确定可靠性系数,从而扩大了粒子群的搜索空间,降低了过往局部最优化的可能性。实验结果表明,FR-MOS 在多目标性能指标方面取得了良好的成效,并且优于 Hu 等提出的 PSO 算法。

Hu 等和 Farid 等的算法尽可能多地选择了可靠性较强

的虚拟机,而 Tang 等^[47]构建了一个多云系统容错 workflow 调度框架,从优化 workflow 执行方案的角度在最小化完工时间和成本的同时提高了 workflow 调度的可靠性。Tang 等将不同的多云供应商计费机制集成到所提出的调度框架中,利用威布尔分布分析任务执行的可靠性和风险率,从而重复执行风险率高的任务。考虑到重复任务会延长虚拟机的使用时间,带来高额的成本,Tang 等提出了一种容错成本高效的 workflow 调度算法(FCWS),用于根据需要调整可靠性和成本的比值参数,从而做出更贴合用户需求的权衡。该方法在提升可靠性方面表现良好,美中不足的是考虑的云提供商的计费机制相对简单,不具有代表性。

3.2.4 其他面向多目标优化的多云 workflow 调度

在其他面向多目标优化的多云 workflow 调度中,当前的主要方法为将能耗、吞吐量、资源利用率等目标与成本、完工时间等目标相结合进行折中调度。

多云环境使得低能耗的虚拟机选择变多,为了更好地利用这一优势,Mishra 等^[48]提出了基于维护 SLA 的多云网络能量感知任务分配算法(ETAMCN)。为了尽可能缩短完工时间,该方法采用递归处理有依赖关系的任务方式,先递归处理父任务,再处理子任务。对于每个任务,先根据 ETC(Expected Time to Complete)矩阵选出满足截止日期的虚拟机集合,然后从该集合中根据 Energy Consumption(EC)矩阵选择最小能耗的虚拟机分配给该任务。该方法虽然没有显著缩短完工时间,但在降低能耗方面有显著成效。

多云环境中现有的调度方法大多适用于处理静态 workflow

中的数据,而对数据量变化较大又需要实时处理的流式工作流却无能为力。为了应对流式工作流中输入数据量突增的场景,Barika 等^[49]建立了一个在成本约束下保持较高的吞吐量的工作流调度机制。该机制根据用户愿意支付的成本上界和工作流数据输入的最大速率为工作流分配虚拟机,当输入流的速度超过约定机制的吞吐量时采用丢弃方案。实验结果表明,该方案在保持吞吐量和减少成本方面具有良好的表现,这为处理输入数据量变化较大的工作流提供了新的思路。

多云环境虽然提供了更多可选的资源,但多数调度算法没有考虑到资源利用率低的问题。针对该问题,Chen 等^[50]设计了一种基于自适应资源分配和整合的在线工作流调度算法(OWS-A2C)。该方法根据多个工作流的性能需求分配和

合并实例:将父任务和子任务相同,但又占据短于一般定价单位时间的并行任务转换成串行任务,并将它们调度到处理速度最快的虚拟机;将没有占据整个定价单位时间的任务尽量合并到同一虚拟机。虚拟机的高度利用减少了工作流在云间传输的损耗,使得这两次优化大幅度提高了资源利用率,降低了执行成本。大量仿真实验证明,该方法能够在多单元集成电路中实现自适应、高效的软件调度。

3.3 多云环境下工作流调度比较

3.1 节和 3.2 节从多云下各类调度目标的角度对多云工作流调度方法做了充分的研究分析。为了更加全面地综合分析和对比,表 2 列出了从方法描述、关键实现、工具、特点等方面对多云调度算法的比较分析。

表 2 多云调度算法的比较分析

Table 2 Comparative analysis of multi-cloud scheduling algorithms

文献	目标	方法描述	关键实现	工具	特点
[32]	成本	提出了缩短虚拟机租赁总时间的成本感知算法	将资源需求量大的任务优先分配给容量大、带宽充裕、出入口成本低的主机	Python	通过缩短虚拟机的租赁时长来降低总成本,具有很好的性能
[34]	完工时间	提出了能够自适应缓存、缓存站点选择和动态工作流调度的算法	该算法必须是缓存感知的,以决定重用缓存数据还是完全重新执行工作流	OpenAlea	能够应对工作负载难以估计的工作流,或者计算能力在执行期间变化较大的环境
[47]	可靠性、完工时间和成本	构建多云系统容错工作流调度框架,利用威布尔分布分析任务执行可靠性和风险率,提出低成本、高容错性的工作流调度算法(FCWS)	威布尔分布分析后,风险率高的任务可被重复执行。FCWS 算法调整可靠性和成本的比值在最终结果中的占比	CloudSim	该方法可靠性强、成本低,但设置的计费机制简单
[48]	完工时间和能耗	提出了一种基于维护 SLA 的多云网络能量感知任务分配算法	根据 ECT 矩阵和 EC 矩阵分别控制完工时间和能耗	CloudSim	降低了能耗,但缩短总完工时间的效果不明显
[50]	资源利用率和成本	提出了一种基于自适应资源分配和整合的 OWS-A2C 方法	将满足条件的并行任务串行,合并任务到同一虚拟机执行	Python	未考虑任务之间的数据传输,也未考虑执行性能波动对结果的干扰
[51]	完工时间和成本	提出了一种自适应粒子群优化(PSO)技术	采用交叉算子和遗传算法作为变异算子,解决了传统粒子群算法存在的早熟收敛问题	64 位 Windows 10, 32 GB 内存	在粒子群算法的基础上提高了种群的多样性,具有更好的性能
[52]	成本和响应时间	提出了一种混合的基于遗传算法的方法	采用修复算法将超预算解的自适应子集转化为预算可行解	Sprint IP Network	可以有效地繁育优良的后代,进而在满足预算条件下缩短了平均响应时间
[53]	成本和能耗	提出了一种能源感知的工作流调度算法 GREENGA	将遗传算法首次应用于能耗和成本权衡问题	IoTSimSDWAN	可将绿色能源用于工业工作流执行的使用量提高,但成本略有增加
[54]	吞吐量、成本	提出了 StreamWS 算法,将吞吐量性价比高的虚拟机优先分配给关键路径上的模块	优先满足关键路径的虚拟机分配和升级	Java	既提高了吞吐量,又保证了关键模块的执行效率

4 未来研究方向

现有的关于多云环境下的工作流调度的研究大部分的优化目标是完工时间或成本,研究内容大多使用不同的方法来降低算法时间复杂度、增加求解多样性,较少考虑一些其他的调度目标或与新技术结合的方法,因此多云工作流调度的研究还有很长的路要走。基于最新相关文献,我们认为多云工作流调度的未来方向和发展趋势还应该包含以下方面。

4.1 不确定性工作流调度

多云环境下弹性的资源提供和虚拟机性能的不稳定性使得工作流的执行面临着诸多不确定性,然而当前调度算法大多默认任务启动/执行/完成时间、任务间数据传输时间等可准确预测,忽视这些不确定因素往往会导致工作流超时,增加

工作流执行的服务租用成本^[55]。

未来需要进一步在多云中研究考虑不确定性的工作流调度。例如,在为工作流分配服务资源时,可以通过控制在服务实例上等待的任务数量来控制不确定性传播^[56]。在设计算法调度目标时,可以将健壮性纳入目标范围,从而生成被资源不确定性影响小的调度解^[57]。

4.2 能耗与其他目标的联合调度优化

多云环境提供了更多可选的低能耗资源,这为降低工作流在云平台上执行产生的能耗提供了新的思路。然而,多云下现有的调度方案多数以牺牲时间或成本的方式来换取低能耗。Sambit 等^[48]提出的基于维护 SLA 的多云网络能量感知任务分配算法(ETAMCN),虽在降低能耗方面有显著成效,但没有显著缩短完工时间。Wen 等^[53]在一组私有 SDWAN

连接的数据中心上调度数据驱动的工业 workflow 应用程序,提出的方案将绿色能源用于 workflow 执行的使用量提高 3 倍,但成本有所增加。无论是对供应商还是对用户而言,以牺牲其他性能来换取低能耗的方案都难以推行。因此,如何在 workflow 调度过程中在降低能耗的同时优化完工时间、成本等其他目标是很有价值的研究方向。

4.3 与边缘服务器协同的调度优化

多云环境在提供了更多可选资源的同时也增加了 workflow 调度的复杂程度,给一些对实时性比较敏感的应用带来了挑战。为了减少网络延迟等因素对数据转移的影响,进而缩短 workflow 调度的响应时间,当前的做法通常为,由全局调度器将 workflow 中的任务模块分段调度到离数据源较近的虚拟机附近^[41,43,54]。为了进一步降低网络成本,未来可以将对实时性要求高的 workflow 交由边缘服务器处理。由边缘服务器为 workflow 分配资源,或将其分配到邻近的边缘服务器、云数据中心^[60]。但由于边缘服务器的计算能力和存储能力有限,对数据规模较大 workflow 的处理时间可能会超过用户对实时性的要求^[61]。因此,如何合理利用边缘服务器的快速响应能力和全局调度器的强大处理能力,让它们协同处理 workflow 调度问题,具有广阔的研究前景。

4.4 虚拟机和 Serverless 平台混合调度

Serverless 平台是一种新的云计算范式,服务器数量及容量等决策由无服务器平台负责,并根据工作负载的需要自动提供服务器容量^[62]。目前所有主要的云提供商都有无服务器的产品,主要以功能即服务(FaaS)的形式提供:亚马逊 Lambda/Azure Functions,谷歌 Cloud Functions 和 IBM Cloud Functions^[63]。随着无服务器技术逐渐成熟,无服务器平台和虚拟机并存已经成为一种必然现象^[64]。然而,当前多云下的 workflow 调度模型大多基于虚拟机进行调度。针对虚拟机和无服务器混合调度能否更好地降低成本、缩短完工时间或降低能耗等问题,有待进一步研究。

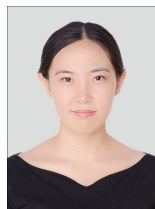
结束语 随着用户对云资源需求的增加,继续使用传统单一的云环境会存在本地资源不足、费用较高等问题。为了减小用户对单个云的依赖,多云环境应运而生。同时,workflow 调度又是多云研究的关键问题之一。本文首先对多云下 workflow 调度技术做了深入的研究,根据具体目标对文献进行分类综述,又将设计的调度方法从优化目标、方法描述、关键实现、工具和特点等维度进行分析和对比。最后,基于已有研究和最新技术发展,讨论了未来研究的方向和趋势。

参 考 文 献

- [1] LEE K. Comments on Secure Data Sharing in Cloud Computing Using Revocable-Storage Identity-Based Encryption [J]. IEEE Transactions on Cloud Computing, 2020, 8(4): 1299-1300.
- [2] CUI J, ZHANG X, ZHONG H, et al. Extensible conditional privacy protection authentication scheme for secure vehicular networks in a multi-cloud environment [J/OL]. IEEE Transactions on Information Forensics and Security, 2019, 15: 1654-1667. <https://ieeexplore.ieee.org/abstract/document/8865553>.
- [3] ZHANG B, ZENG Z, SHI X, et al. A novel cooperative resource provisioning strategy for Multi-Cloud load balancing [J/OL]. Journal of Parallel and Distributed Computing, 2021, 152: 98-107. <https://www.sciencedirect.com/science/article/abs/pii/S0743731521000241>.
- [4] Multicloud. [EB/OL]. (2021-08-09). [2021-11-08]. <https://www.ibm.com/cloud/learn/multicloud>.
- [5] HEILIG L, LALLA-RUIZ E, VOß S. Modeling and solving cloud service purchasing in multi-cloud environments [J/OL]. Expert Systems with Applications, 2020, 147: 113165. <https://www.sciencedirect.com/science/article/abs/pii/S0957417419308826>.
- [6] GHARAMANI M H, ZHOU M C, HON C T. Toward cloud computing QoS architecture: Analysis of cloud systems and cloud services [J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(1): 6-18.
- [7] KANG S, VEERAVALLI B, AUNG K M M. Dynamic scheduling strategy with efficient node availability prediction for handling divisible loads in multi-cloud systems [J/OL]. Journal of Parallel and Distributed Computing, 2018, 113: 1-16. <https://www.sciencedirect.com/science/article/abs/pii/S0743731517302824>.
- [8] Global Cloud Computing Market Size & Share Will Reach USD 1025.9 Billion by 2026: Facts & Factors [EB/OL]. (2021-01-22) [2021-02-28]. <https://www.globenewswire.com/news>.
- [9] IRANMANESH A, NAJI H R. DCHG-TS: a deadline-constrained and cost-effective hybrid genetic algorithm for scientific workflow scheduling in cloud computing [J]. Cluster Computing, 2021, 24(2): 667-681.
- [10] TOMARCHIO O, CALCATERA D, DI MODICA G. Cloud resource orchestration in the multi-cloud landscape: a systematic review of existing frameworks [J]. Journal of Cloud Computing, 2020, 9(1): 1-24.
- [11] HOSSEINZADEH M, GHAFOR M Y, HAMA H K, et al. Multi-objective task and workflow scheduling approaches in cloud computing: a comprehensive review [J/OL]. Journal of Grid Computing, 2020: 1-30. <https://link.springer.com/article/10.1007/s10723-020-09533-z>.
- [12] ARUNARANI A R, MANJULA D, SUGUMARAN V. Task scheduling techniques in cloud computing: A literature survey [J/OL]. Future Generation Computer Systems, 2019, 91: 407-415. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X17321519>.
- [13] KUMAR M, SHARMA S C, GOEL A, et al. A comprehensive survey for scheduling techniques in cloud computing [J/OL]. Journal of Network and Computer Applications, 2019, 143: 1-33. <https://www.sciencedirect.com/science/article/abs/pii/S1084804519302036>.
- [14] MASDARI M, ZANGAKANI M. Efficient task and workflow scheduling in inter-cloud environments: challenges and opportunities [J]. The Journal of Supercomputing, 2020, 76(1): 499-535.
- [15] ALALUNA M, VIAL E, NEVES N, et al. Secure multi-cloud network virtualization [J/OL]. Computer Networks, 2019, 161: 45-60. <https://www.sciencedirect.com/science/article/abs/pii/S1389128618312155>.
- [16] WANG G, NG T S E. The impact of virtualization on network performance of amazon ec2 data center [C] // 2010 Proceedings

- IEEE INFOCOM. IEEE, San Diego, CA, USA, 2010; 1-9.
- [17] REDKAR T, GUIDICI T, MEISTER T. Windows azure platform[M]. New York: Apress, 2009.
- [18] ZHU Q H, TANG H, HUANG J J, et al. Task Scheduling for Multi-Cloud Computing Subject to Security and Reliability Constraints [J]. IEEE/CAA Journal of Automatica Sinica, 2021, 8(4): 848-865.
- [19] LIU J, REN J, DAI W, et al. Online Multi-Workflow Scheduling under Uncertain Task Execution Time in IaaS Clouds [J/OL]. IEEE Transactions on Cloud Computing, 2019; 1-1. <https://ieeexplore.ieee.org/abstract/document/8669862>.
- [20] CHEN W, XIE G, LI R, et al. Efficient task scheduling for budget constrained parallel applications on heterogeneous cloud computing systems [J/OL]. Future Generation Computer Systems, 2017, 74: 1-11. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X16304411>.
- [21] CAO B, ZHAO J, GU Y, et al. Security-Aware Industrial Wireless Sensor Network Deployment Optimization[J]. IEEE Transactions on Industrial Informatics, 2020, 16(8): 5309-5316.
- [22] ROSA M J F, RALHA C G, HOLANDA M, et al. Computational resource and cost prediction service for scientific workflows in federated clouds[J/OL]. Future Generation Computer Systems, 2021, 125: 844-858. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X21002922>.
- [23] PAKNEJAD P, KHORSAND R, RAMEZANPOUR M. Chaotic improved PICEA-g-based multi-objective optimization for workflow scheduling in cloud environment [J/OL]. Future Generation Computer Systems, 2021, 117: 12-28. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X20330260>.
- [24] FAKHFAKH F, KACEM H H, KACEM A H. Workflow scheduling in cloud computing: a survey[C]// 2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations. IEEE, Ulm, Germany, 2014: 372-378.
- [25] MING T, OTA K, DONG M X. DSARP: dependable scheduling with active replica placement for workflow applications in cloud computing [J/OL]. IEEE Transactions on Cloud Computing, 2016; 1069-1078. <https://ieeexplore.ieee.org/abstract/document/7742980>.
- [26] MOHAMMADI S, POURKARIMI L, PEDRAM H. Integer linear programming-based multi-objective scheduling for scientific workflows in multi-cloud environments[J]. The Journal of Supercomputing, 2019, 75(10): 6683-6709.
- [27] ZHOU J L, WANG T, CONG P J, et al. Cost and makespan-aware workflow scheduling in hybrid clouds [J/OL]. Journal of Systems Architecture, 2019, 100: 101631. <https://www.sciencedirect.com/science/article/pii/S1383762119302954>.
- [28] BUGINGO E, ZHANG D, CHEN Z, et al. Towards decomposition based multi-objective workflow scheduling for big data processing in clouds [J]. Cluster Computing, 2021, 24(1): 115-139.
- [29] GU Y, BUDATI C. Energy-aware workflow scheduling and optimization in clouds using bat algorithm [J/OL]. Future Generation Computer Systems, 2020, 113: 106-112. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X19317066>.
- [30] ZHANG L, LI K, LI C, et al. Bi-objective workflow scheduling of the energy consumption and reliability in heterogeneous computing systems [J/OL]. Information Sciences, 2017, 379: 241-256. <https://www.sciencedirect.com/science/article/abs/pii/S0020025516305722>.
- [31] JIANG F, FERRITER K, CASTILLO C. PIVOT: Cost-Aware Scheduling of Data-Intensive Applications in a Cloud-Agnostic System [R/OL]. Technical Report TR-19-02. RENCI, University of North Carolina at Chapel Hill, 8 pages. <https://renci.org/technical-reports/tr-19-02>, 2019.
- [32] JIANG F, FERRITER K, CASTILLO C. A cloud-agnostic framework to enable cost-aware scheduling of applications in a multi-cloud environment[C]// 2020 IEEE/IFIP Network Operations and Management Symposium (NOMS 2020). IEEE, Budapest, Hungary, 2020; 1-9.
- [33] XU S J, LIU G G. A Novel Scheduling Algorithm for Scientific Workflow in Multi-cloud Environment [J]. Journal of Minjiang University, 2018, 39(5): 63-72.
- [34] HEIDSIECK G, OLIVEIRA D D, PACITTI E, et al. Cache-aware scheduling of scientific workflows in a multisite cloud [J/OL]. Future Generation Computer Systems, 2021, 122(7). <https://www.sciencedirect.com/science/article/abs/pii/S0167739X21000923>.
- [35] LIU J, PACITTIE, VALDURIE Z P, et al. Multi-objective scheduling of Scientific Workflows in multisite clouds [J]. Future Generation Computer Systems, 2016, 63(oct.): 76-95.
- [36] SENTURK I F, BALAKRISHNAN P, ABU-DOLEH A, et al. A resource provisioning framework for bioinformatics applications in multi-cloud environments [J/OL]. Future Generation Computer Systems, 2018, 78: 379-391. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X16301911>.
- [37] ULABEDIN Z, NAZIR B. Replication and data management-based workflow scheduling algorithm for multi-cloud data centre platform [J/OL]. The Journal of Supercomputing, 2021; 1-30. <https://link.springer.com/article/10.1007/s11227-020-03541-2>.
- [38] TAO S, MA H, CHEN G. A genetic-based approach to location-aware cloud service brokering in multi-cloud environment[C]// 2019 IEEE International Conference on Services Computing (SCC). IEEE, 2019.
- [39] MA H, DA SILVA A S, KUANG W. NSGA-II with local search for multi-objective application deployment in multi-cloud[C]// 2019 IEEE Congress on Evolutionary Computation (CEC). IEEE, Wellington, New Zealand, 2019; 2800-2807.
- [40] ZHANG Q. Knowledge Incorporation in Evolutionary Computation [Book Review][J]. IEEE Computational Intelligence Magazine, 2006, 1(4): 58-59.
- [41] SHI T, MA H, CHEN G. Seeding-Based Multi-Objective Evolutionary Algorithms for Multi-Cloud Composite Applications Deployment[C]// 2020 IEEE International Conference on Services Computing (SCC). IEEE, Beijing, China, 2020; 240-247.
- [42] POOLA D, RAMAMOCHANARAO K, BUYYA R. Enhancing reliability of workflow execution using task replication and spot instances [J]. ACM Transactions on Autonomous and Adaptive Systems (TAAS), 2016, 10(4): 1-21.
- [43] SINGH A, CHATTERJEE K. Cloud security issues and challenges: A survey [J/OL]. Journal of Network and Computer Appli-

- cations, 2017, 79: 88-115. <https://www.sciencedirect.com/science/article/abs/pii/S1084804516302983>.
- [44] HU H, LI Z, HU H, et al. Multi-objective scheduling for scientific workflow in multicloud environment [J/OL]. *Journal of Network and Computer Applications*, 2018, 114: 108-122. <https://www.sciencedirect.com/science/article/abs/pii/S1084804518301152>.
- [45] FARID M, LATIP R, HUSSIN M, et al. Scheduling Scientific Workflow Using Multi-Objective Algorithm With Fuzzy Resource Utilization in Multi-Cloud Environment [J/OL]. *IEEE Access*, 2020. <https://ieeexplore.ieee.org/abstract/document/8976136>.
- [46] MENDELJ M. Fuzzy logic systems for engineering : a tutorial [J]. *Proceedings of the IEEE*, 1995, 83(3): 345-377.
- [47] TANG X. Reliability-Aware Cost-Efficient Scientific Workflows Scheduling Strategy on Multi-Cloud Systems [J/OL]. *IEEE Transactions on Cloud Computing*, 2021. <https://ieeexplore.ieee.org/abstract/document/9349203>.
- [48] MISHRA S K, MISHRA S, ALSAYAT A, et al. Energy-aware task allocation for multi-cloud networks [J/OL]. *IEEE Access*, 2020, 8: 178825-178834. <https://ieeexplore.ieee.org/abstract/document/9205994>.
- [49] BARIKA M, GARG S, CHAN A, et al. Scheduling algorithms for efficient execution of stream workflow applications in multi-cloud environments [J]. *IEEE Transactions on Services Computing*, 2022, 15(2): 860-875.
- [50] CHEN Z, LIN K, LIN B, et al. Adaptive Resource Allocation and Consolidation for Scientific Workflow Scheduling in Multi-Cloud Environments [J/OL]. *IEEE Access*, 2020, 8: 190173-190183. <https://ieeexplore.ieee.org/document/9233315>.
- [51] MANAM S, MOESSNER K, VURAL S. A Cost-Aware Strategy for Deadline Constrained Scientific Workflows [J]. *Journal of Physics: Conference Series*, 2020, 1577(1): 012036.
- [52] SHI T, MA H, CHEN G, et al. Location-aware and budget-constrained service deployment for composite applications in multi-cloud environment [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2020, 31(8): 1954-1969.
- [53] WEN Z, GARG S, AUJLA G S, et al. Running Industrial Workflow Applications in a Software-Defined Multicloud Environment Using Green Energy Aware Scheduling Algorithm [J]. *IEEE Transactions on Industrial Informatics*, 2020, 17(8): 5645-5656.
- [54] LI R, WU C Q, HOU A, et al. On scheduling of high-throughput scientific workflows under budget constraints in multi-cloud environments [C] // 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCLOUD/SocialCom/SustainCom). *IEEE*, Melbourne, VIC, Australia, 2018: 1087-1094.
- [55] CALZAROSSA M C, DELLA VEDOVA M L, MASSARI L, et al. Multi-Objective Optimization of Deadline and Budget-Aware Workflow Scheduling in Uncertain Clouds [J/OL]. *IEEE Access*, 2021, 9: 89891-89905. <https://ieeexplore.ieee.org/abstract/document/9461757>.
- [56] CHEN H, ZHU X, LIU G, et al. Uncertainty-Aware Online Scheduling for Real-Time Workflows in Cloud Service Environment [J/OL]. *IEEE Transactions on Services Computing*, 2018. <https://ieeexplore.ieee.org/abstract/document/8443134>.
- [57] ZHANG X D. Multi-objective optimization of workflow scheduling in uncertain cloud environment [J/OL]. *Computer Engineering and Design*, 2021, 42(7): 9. <https://www.cnki.com.cn/Article/CJFDTOTAL-SJSJ202107021.htm>.
- [58] XIE Y, ZHU Y, WANG Y, et al. A novel directional and non-local-convergent particle swarm optimization based workflow scheduling in cloud-edge environment [J/OL]. *Future Generation Computer Systems*, 2019, 97: 361-378. <https://www.sciencedirect.com/science/article/abs/pii/S0167739X18321332>.
- [59] XU X, CAO H, GENG Q, et al. Dynamic resource provisioning for workflow scheduling under uncertainty in edge computing environment [J/OL]. *Concurrency and Computation: Practice and Experience*, 2020. <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5674>.
- [60] DAS A, LEAF A, VARELA C A, et al. Skedulix: Hybrid Cloud Scheduling for Cost-Efficient Execution of Serverless Applications [C] // 2020 IEEE 13th International Conference on Cloud Computing (CLOUD). *IEEE*, Beijing, China, 2020: 609-618.
- [61] BAARZI A F, KESIDIS G, JOE-WONG C, et al. On Merits and Viability of Multi-Cloud Serverless [C] // Proceedings of the ACM Symposium on Cloud Computing. Seattle, WA, USA, 2021: 600-608.
- [62] MA Z H, LIU B, LIN W W, et al. A Survey of resource scheduling for serverless platform [J]. *Computer Science*, 2021, 48(4): 261-267.



YU Hao-wen, born in 1997, postgraduate. Her main research interests include cloud computing and resource scheduling.



LIU Bo, born in 1968, Ph.D, professor, is a member of China Computer Federation. His main research interests include cloud computing, big data technology and distributed security technology.