

云计算环境下调度算法的趋势分析

吴煜祺¹, 曾国荪¹, 曾 媛²

(1 同济大学 计算机科学与技术系, 上海 200092; 2 上海理工大学 管理学院, 上海 200093)

摘 要: 云资源管理是云计算成败的关键, 而调度作为云资源管理的重要手段, 直接影响到云计算的效果。但是, 云计算的异构性与动态性, 导致云计算环境下的调度研究复杂和困难。因此, 在传统并行分布调度工作的基础上, 文中采用五要素刻画调度问题的本质, 并且给出每个要素的具体表现形式。通过分析云调度的现有研究成果, 结合云计算存在的困难和挑战, 指出云计算中调度算法的发展趋势, 为将来的云调度研究指明方向和思路。

关键词: 云计算; 调度问题; 算法综述; 趋势分析

中图分类号: TP338

文献标识码: A

文章编号: 1000-7180(2012)09-0103-06

Trend Analysis for Scheduling Algorithm in Cloud Computing

WU Yu-qi¹, ZENG Guo-sun¹, ZENG Yuan²

(1 Department of Computer Science and Technology, Tongji University, Shanghai 200092, China;

2 School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Resource management plays an important role in cloud computing, while scheduling is one of the key techniques in resource management, which has a direct influence on cloud computing. However, due to the heterogeneous and dynamic features in cloud computing, cloud scheduling is complicated and difficult. This paper uses five factors to describe general scheduling problems, and gives examples of different types for each factor. We analyze the traditional scheduling algorithms in parallel and distributed environment, and the current research in cloud computing. And we give the trend for cloud scheduling under its difficulties and challenge. Our work indicates the directions and ideas for future research in cloud scheduling.

Key words: cloud computing; scheduling problem; algorithm survey; trend analysis

1 引言

云计算是在并行计算、分布式计算、集群计算、网格计算的基础上发展起来的一种新兴计算模式, 它除了拥有上述计算模式所有的特点之外, 还具有“按需、高效、可靠、低价”的优势, 并且正在广为接受和认可。

云计算是人类计算追求的梦想, 梦想变成现实取决于高效可靠的云资源管理系统, 即云平台的构建和管理是关键。资源管理是云计算的核心内容, 目

的是屏蔽底层资源的异构性和复杂性, 把分散的各种各样的云计算资源管理起来, 以达到资源的有效共享和协同工作, 提高资源的利用率。并且为用户任务按需分配合理的云计算资源, 实现负载平衡。而调度作为资源管理系统最重要的手段, 影响到云资源管理的效果。科学的调度不仅可以大大提高云环境的资源利用率, 也可以充分开发资源的异构特性, 彰显资源个性, 物尽其用。不但如此, 通过调度, 也可以有效应对云环境中资源进入退出、节点失效、资源故障这些突发事件, 实现事故容错, 提高系统健壮性,

收稿日期: 2012-02-10; **修回日期:** 2012-03-20

基金项目: 国家自然科学基金项目(61103068); 国家“八六三”计划项目(2009AA012201); NSFC 微软亚洲研究院联合资助项目(6097155); 教育部网络时代的科技论文快速共享专项研究课题(20110740001); 教育部博士点基金项目(20090072110035); 上海市优秀学科带头人计划项目(10XD1404400)

保证服务质量。

为了应对复杂多变的云计算环境,调度问题的研究就充满了挑战,把握云计算环境下的调度方向就显得非常重要。因此本文在研究传统并行分布调度工作的基础上,对调度问题进行本质分析,然后结合云调度中已有的工作和云计算的挑战困难,指出云计算环境下调度算法的发展趋势。

2 调度问题分析

一般来说,调度就是按照某种策略,将任务合理地分配到处理单元上执行,以达到完成时间最短、提高系统利用率等目的。调度是实现高性能计算的关键技术之一。经过研究和总结,调度问题包含五个要素,可以用五元组来表示 $S=(T, R, G, C, P)$, 其中 T 代表计算任务, R 代表执行的资源环境, G 代表调度目标, C 代表约束条件, P 代表调度策略。调度的本质就是通过有效的 P , 将 T 在 R 中调度执行,同时满足 C , 达到 G 。通过对调度论文的广泛阅读,总结出针对不同的应用场景,五要素的不同表现形式。

2.1 调度五要素

(1) 计算任务 T : 是指满足用户需求和功能的计算机程序、进程、作业或服务。用户的需求千差万别,计算任务就存在各种类型。根据任务的耦合程度可分为:独立任务和可划分任务。根据任务之间的相关性可分为:元任务和依赖任务,其中依赖任务通常以 DAG 图表示。在考虑相关性的基础上,加上任务的规模这一要素,可将任务分为:元任务池和工作流。根据任务对资源要求的强度特征可分为:计算密集、数据密集、通信密集、I/O 密集任务。根据服务性质可分为:科学计算、事务处理、实时处理任务。根据到达规律可分为:均匀、周期、随机、成批任务。根据优先级可分为:抢占、非抢占任务等。

(2) 执行资源环境 R : 是指计算任务的运行平台,由处理器、内存、I/O 设备、网络等硬件资源以及软件支持系统组成。根据计算模式可分为:单机系统、多机系统。多机系统又可分为:并行计算、分布计算、集群计算、网格计算、移动计算、普适计算、P2P 计算、云计算系统等。其中,并行计算系统根据体系结构又可分为:SIMD、MIMD、PVP、SMP、MPP、COW、DSM 等。根据类型和性质可分为:同构系统和异构系统。其中本文所要研究的云计算环境属于异构、分布的多机系统。

(3) 调度目标 G : 是调度过程中最高、或特定、或焦点的要求。根据不同需求、不同的执行环境、不

同的场合,存在许多调度目标。来自用户方面的调度目标有:最短完成时间、截止期限、服务成本、正确性、公平性、安全性、可靠性等。来自服务器或资源厂商方面的调度目标有:资源利用率、收益、运行成本、吞吐量、容错、负载均衡、声誉、能耗等。来自调度器本身的调度目标有:响应速度、效率、成功率、稳态、智能等。

(4) 调度约束 C : 是调度过程中的一般要求,即必须满足的限制条件,为调度决策合理性的制约。调度约束表现为一种条件关系,限定过分地追求特定调度目标,保证调度的合理执行。调度目标与约束之间存在复杂关系,当要求的强烈程度或等级发生变化时,目标可能变成约束,约束可能变成目标。例如调度目标 T 为最小完成时间时,成本和资源规模就可能为调度约束 C 。但是如果调度目标为最小成本时,则完成时间和功耗就可能变成调度约束 C 。

(5) 调度策略 P : 侧重于描述调度的具体实现方案,是调度算法形成的思想基础。根据调度策略决策的时机,调度可分为:静态和动态调度。根据考虑的系统规模,调度可分为:局部和全局调度。根据调度的执行方的不同,调度可分为:他调度和自调度。根据调度策略是否随系统状态信息变化而自动调整,调度可分为:自适应和非自适应调度。根据调度程序执行调度操作时是否允许任务抢占执行,调度可分为:抢占式和非抢占式调度。根据任务是否允许迁移执行,调度可分为:迁移和非迁移调度。考虑调度算法在什么地方执行、调度信息存储在什么地方、以及调度算法所使用的技术到底有多复杂,可以将调度方式分为中心式调度、层次式调度和分布式调度。另外调度策略中还包含了启发式方法,可分为:表调度、分簇调度、复制调度和非确定性调度等。

2.2 调度的公式化描述

为了本质地刻画调度问题内涵,可以引入变量、参数、概率等数学符号,通过方程组来刻画调度各要素之间相互关系,建立数学模型。例如,对于调度目标,可以用最大值最小值函数 $\max(y)$, $\min(y)$ 表示;多目标可用单目标函数的方程组表示,也可用目标多元函数表示。对于调度约束,可采用:(1)区间表示法,例如 $x \in [a, b]$; (2)等式或不等式表示法,例如 $f(t) \leq T$; (3)物理公式表示法,例如系统功耗

可表示为 $\sum_{i=1}^n \sum_{j=1}^m x_{ij} t_{ij} p_i$; (4)概率表示法,例如计算

节点的可靠性可以表示为 $\prod_{i=1}^n r_i(\theta_i)$.

调度问题的数学公式化描述就是由调度目标函数同多个约束条件联立起来的方程组. 例如:当用户希望得到高效服务,而服务提供商希望获得高利润,同时还要求系统的能耗控制在一定范围内. 此时,可以建立最小完成时间、最大利润的数学模型,如下面的方程组所示. 其中, z_1 为任务完成时间, z_2 为利润,相应的约束条件为每台设备的功耗控制在一定范围内.

$$\begin{cases} \min(z_1) = \sum_{i=1}^n \sum_{j=1}^m x_{ij} t_{ij} \\ \max(z_2) = \sum_{i=1}^n \sum_{j=1}^m x_{ij} (prof)_{ij} \\ \sum_{i=1}^n \sum_{j=1}^m x_{ij} t_{ij} p_i \in [W_{\min}, W_{\max}] \\ \sum_{i=1}^n x_{ij} \leq 1 \end{cases}$$

2.3 调度算法的有效性评价

自调度问题提出至今,产生了大量调度算法. Kwok 等^[1]提出了并行环境中调度算法的评价体系,即从以下四个方面评价调度算法的有效性:

(1) 精确性(Precision):一个调度算法必须具有高精度以及鲁棒性. 一般来说,调度是 NP 难题,设计的调度算法只能寻找到近优解,因此获得的解精度越高说明调度算法越有效. 例如当调度的目标为完成时间最短,若采用最优调度算法 S_{best} ,任务的完成时间为 t_{best} ;采用可行调度算法 S_{fea} ,取得任务的完成时间为 t_{fea} ,则 S_{fea} 的精确性定义为 $\lambda = (t_{\text{fea}} - t_{\text{best}}) / t_{\text{best}}$, $\lambda \geq 0$. λ 越趋近于 0,说明 S_{fea} 的精度越高,调度算法越优.

(2) 时间复杂度(Time-complexity):一个算法的好坏可以根据时间复杂度评判. 同样,调度算法也可用时间复杂度进行评判. 例如在实时计算任务应用中,一个快速的算法对于有效求解具有重要意义. 例如利用关键路径进行表调度的两个算法 MCP 和 DCP 具有不同的时间复杂度, MCP 算法的时间复杂度为 $O(n^2 \log n)$,而作为动态调度算法的 DCP 的时间复杂度为 $O(n^3)$,其中 n 为任务的数量. 显然, MCP 算法比 DCP 算法更优.

(3) 可扩展性(Scalability):对于运行在处理机上的某个调度算法,当处理机数目增大时,若同时增大一定的工作量能维持整个并行计算系统的平均速度保持不变,则称此调度算法是可扩展的. 设 P, P'

分别是并行系统扩展前后中处理机的个数, $\bar{v}, \bar{v}'$ 分别是并行系统扩展前后的平均速度, W, W' 分别是并行系统扩展前后计算任务的工作负载,则记 $\Psi(P, P') = \frac{\bar{v}}{\bar{v}'}$ 为等速度可扩展性函数. 特别在同构

并行计算系统中, $\Psi(P, P') = \frac{W/P}{W'/P'} = \frac{WP'}{W'P}$ 为等速度可扩展性函数. $\Psi(P, P')$ 的值介于 0 与 1 之间,值越大表示调度算法的可扩展性越好.

(4) 可用性(Applicability):任何算法的目的都是为了应用,所以调度算法能够真正应用于实际的环境中,是调度算法的一个必要要求. 为了达成这个目标,在设计调度算法时,不能过于理论化和理想化,必须考虑实际环境和复杂因素,例如计算量、通信量、链路冲突、处理机间的互连结构等.

3 云计算环境下调度算法综述

云计算环境不同于以往的调度执行环境. 目前,国内外学者针对云环境的特征和要求,提出不同的调度算法. Huang 等^[2]针对云环境资源分布的特点,提出了动态分布式调度算法 CASA,该算法中每个调度节点都拥有一个元调度器,用来接收来自本地用户所提交的任务并且进行任务分配,同时各个调度节点之间也保持信息共享,以达到节点间负载均衡的目的. Lee 等^[3]就云环境中扩大资源共享所带来的矛盾,即通过提高资源利用率扩大了云服务提供商的经济利益,但却降低了用户要求的性能水平,例如在完成时间内无法完成任务的这对矛盾,设计了一种新的计算收益的模型,此模型不光考虑了当前服务所带来的收益,同时也考虑了其它任务所能带来的收益,并提出了一个新的服务请求算法,使得在用户完成时间要求的约束下,达到最大化服务提供商利益的目标.

Abrishami 等^[4]为了解决云环境中工作流在用户给定截止期前完成的约束下,最小化工作流的总执行开销的目标的调度问题,提出了 SC-PCP(SaaS Cloud Partial Critical Paths)算法. 此调度算法在调度目标的指导下,首先调度工作流中的关键路径上的节点,在关键节点调度完毕后,将每个关键节点作为出口节点,将每个关键节点的开始时间作为其父节点的截止期,再接着寻找部分关键路径,直到所有的节点调度完毕. Garg 等^[5]综合考虑了环境可持续和经济可持续之间的关系,以最小碳排放量和最大化经济效益为调度目标,提出了一种基于碳/能量的

云环境多数据中心调度算法,在每个数据中心上都有一个调度器,调度器可以提供碳排放率、性能系数 COP(Coefficient of Performance)、电价、CPU 供电频率、CPU 最大频率下的 MIPS(Million Instructions Per Second)和 CPU 工作频率范围,并且根据这些因素给出每个数据中心的效益和碳排放量的表达式. Mezmaz 等^[6]对云环境下的依赖任务,以最小完成时间和最低能耗为调度目标,提出了基于遗传算法和能耗感知的混合调度算法,其中能耗优化通过采用电压动态调整的技术实现.

Lucas-Simarro 等^[7]对多个价格动态变化的云环境,设计了云代理结构,利用调度达到最优服务. 所设计的代理结构分为中央数据库、云管理器、调度器和虚拟机管理器,其中它着重阐述了调度器的运行机理,给出了服务中开销最小和性能最优这两个参数的公式,以及各种约束的表达式,例如性能约束、开销约束、配置约束、实例类型约束、再分配约束、利用率约束等,这样调度就可以按照不同的服务参数和用户所给出的约束组合,给出调度策略. Xu 等^[8]根据市场经济模型,对云环境中资源提供方和用户之间的关系进行分析,采用伯格模型(Berger Model),对云中资源、用户任务、资源分配和公平性评价公式进行建模,其中用户任务按照完成时间和带宽需求进行分类,并且给出了两种公平性的约束评价公式. Riakiotakis 等^[9]为了解决了调度中出现的实时估计任务的预计完成时间过长,使得等待任务的预计开始时间推迟,从而造成任务的响应时间过长的问題,提出了一种云环境下的独立任务集的协调重调度算法,该算法通过等待队列上的任务进行重调度,使等待任务在预计开始时间前执行.

4 云计算调度算法趋势分析

4.1 云计算中存在的挑战问题

目前,云计算正在经历从理想概念提出到关键技术实现的转变,从理论体系架构研究到实际应用平台构建的转变,从云解决方案到企业效益盈利的转变. 但是云计算要想完全实现从学术研究到产业化的转变,还面临着各种问题和挑战:(1)弹性资源管理. 云环境中硬件资源数量众多、种类异构,用户服务任务复杂多样,需求模式未知不定,导致现有的资源管理策略很难满足各种要求.(2)按需服务. 云计算的一个重要特征是能够根据用户的服务请求,按需提供个性化的服务. 但是,服务的提供者如何能够实现这一目标,目前还没有成熟的技术方案.(3)

资源的可用性. 在传统的网络环境中,经常会出现服务器异常宕机、服务被攻击、操作系统崩溃等问题,这些问题在云计算环境中依然存在.(4)虚拟机迁移. 为了保证云计算环境的高鲁棒性,需要快速响应能力,但在进行负载热点检测和虚拟机迁移初始化时,现有技术不能根据负载的变化而快速的做出反应. 另外,当虚拟机迁移时,内存的状态必须和资源一起,连续、有效地进行迁移,这是较难做到的.(5)能耗管理. 在云计算数据中心,巨大的电能开销带来了沉重的经济负担,然而任何节能技术都是以影响性能为代价,所以如何根据不同场景,选择能耗和性能的权衡,就成了亟待解决的问题.(6)数据存储. 云计算环境下的软件框架通常是基于互联网的文件系统,例如 GFS 和 HDFS,这些文件系统和传统分布式文件系统在存储结构、访问模式、应用编程接口等方面都不同. 更重要的是要让用户有信心将自己的数据愿意存储在数据中心,并且数据保持完整可靠,不出问题.(7)数据传输. 云环境中,数据传输服务是最主要的服务类型之一,并且云服务主要是通过互联网的方式提供给用户. 因此,高带宽,低延迟,无错误的数据传输是保证云计算得以推广的基本条件. 但是,在现有互联网通信条件下,缓慢的网速远远不能满足大量数据传输的需要.(8)流量监管. 对于云计算数据中心来说,对数据流量的管理和分析不仅可以用来改善用户的操作体验,还可以用来对整个网络进行管理和规划. 但是,要想把现有的针对互联网的流量测量和分析方法扩展到云数据中心,还存在很多问题.(9)安全隐私. 在云计算环境中,运维方必须防止数据中心的数据被非法截取、攻击、篡改. 对于用户来说,担心个人的隐私信息被未授权的看到. 可见如何根据云计算环境的特征,设计安全保密策略,是云计算面临的一大难题.(10)标准化工作. 随着云计算的发展,人们可通过各种不同的智能终端来提交服务请求,这些不同的终端设备都需要与云计算平台交换和共享数据. 但是,云计算平台如何提供统一的接口来实现服务接入,如何实现多个云平台之间的互联互通,即如何实现标准化,同样是云计算面临的一大挑战.

4.2 云调度的发展趋势

从以上这些挑战中可以看出,云计算中的问题主要来自于:(1)原有计算模式下的方法不能直接应用于云计算模式中,例如数据管理和流量监管等;(2)云计算环境下产生了许多新需求,例如弹性、按需、成本等,对应新方法的研究尚处于探索状态;(3)

大规模、异构、动态等特征使云环境更加复杂.因此,云资源管理,以及云计算环境下的调度方案都应该根据挑战而进行调整,适应云环境的新特征.按照此思路,本文提出了云调度未来可能的研究和发展趋势:

(1) 保证基本 QoS 的按需调度:云计算环境下基本的 QoS 包括最短完成时间、截止时间、传输速率、服务成本、正确性等.在服务质量的基础上,通过按需调度,用户按照不同的服务质量需求使用云中的资源,如日常生活中的水电煤服务一样,并按实际使用情况付费,完成用户的个性化服务,这就是云计算中的按需服务思想.

(2) 能耗优化调度:云计算在不断追求计算高性能以获得强大计算能力的过程中,大量的能量消耗已成为云计算发展亟待解决的问题.能耗优化除了可以从硬件方面进行改善,也可以利用管理和调度手段进行控制.例如调度可以进行电压动态调整、关闭/休眠低使用率的处理机,从而达到节能的目的.

(3) 多租户并发随机调度:多租户作为云计算的一项技术,使得大量的用户能够共享同一堆栈的软、硬件资源.多租户向云平台请求的服务具有随机性,即服务请求的数量和时机是任意的,可能是单个到达或是成批到达,也可能是均匀、周期、按一定分布到达,这样很可能会导致任务的波动涌现.采用调度技术对服务请求进行动态调整,在保障各租户性能指标的同时,提高整体资源利用率,降低了成本,用来达到租户要求和提供商利益的平衡.

(4) 满足 SLA 预约的供求双赢调度:云服务采用服务等级协议形式,双方的承诺将以协议的形式确立.协议确定了服务双方的义务,精确的定义了服务质量涉及的各项指标,规定了服务双方履行协议的要求和违背协议所承担的责任,保障了资源提供者与使用者之间合作的正常进行.而且,协议签订后,服务内容可以预知,服务方可以根据服务申请进行规划.基于 SLA 的调度既能向用户提供满足协议约束的服务,又可以避免资源使用高峰,合理安排资源使用顺序,促使服务以合理的价格与成本完成交易.

(5) 虚拟机迁移调度:虚拟机在线迁移能够在不停止服务运行的情况下,在极短的时间内,将虚拟机从源宿主机节点迁移到目标节点,以实现节点间资源的负载均衡.该方法是保持虚拟机性能和提高服务可用性的有效方法.一个好的虚拟机迁移调度

方案,不光可以完成指导虚拟机迁移的工作,还可以保障云服务的质量.

(6) 多机故障容错调度:云计算环境是一个开放性的平台,资源数量巨大、种类繁多、广域分布、可动态的加入和退出,导致突发事件、节点失效是不可避免的.如果利用云计算系统构建的核电站、飞行控制系统、交通指挥调度系统等,没有实时故障检测,没有容错备份调度处理方案,往往会引起灾难性的后果.传统集群调度中通常假设一个处理机失效,这对于研究云计算的容错问题显然存在很大的局限性,因此研究多处理机故障容错调度更具有现实意义.

(7) 自服务隐代理分层式群调度:云平台提供的按需服务,一切隐蔽在后台,冥冥当中靠群体协同调度完成.大规模异构云计算环境,调度目标和约束很多,导致各种各样的调度方案,并且每一种调度方案都有其适用和有效的调度任务和资源环境.显然,固定单一集中式的调度方案将成为调度决策瓶颈,顾此失彼,调度效率低下.由此,呼吁个性化的自主调度机制,只有通过自助式服务,才能满足用户的真正要求.

5 结束语

本文通过分析已有调度工作,采用调度五要素用以全面描述调度问题的本质和内涵,并且举例说明了每个要素的具体表现形式.然后对传统并行分布计算中的经典调度算法进行了综合分析,同时通过分析当前云计算环境下的调度研究,结合云计算所带来的新问题和新的挑战,指出了云调度的发展趋势.以上工作不仅可以帮助快速了解并行分布计算中的经典调度算法和结论,更可以对云计算中调度问题的研究起到一定的指导和启发作用,从而进一步推动云计算的研究和应用发展.

参考文献:

- [1] Kwok Y K, Ahmd I. Static scheduling algorithms for allocating directed task graphs to multiprocessors[J]. ACM Computing Surveys, 1999, 31(4):406-471.
- [2] Huang Y, Bessis N, Norrington P, et al. Exploring decentralized dynamic scheduling for grids and clouds using the community-aware scheduling algorithm[J]. Future Generation Computer Systems, 2011, doi: 10.1016/j.future.2011(5):6.
- [3] Lee Y C, Wang C, Zomaya A Y, et al. Profit-driven scheduling for cloud services with data access aware-

- ness[J]. Journal of Parallel and Distributed Computing, 2012, 74(4): 591-602.
- [4] Abrishami S, Naghibzadeh M. Deadline-constrained workflow scheduling in software as a service cloud[J]. Scientia Iranica, 2011(11):47.
- [5] Garg S K, Yeo C S, Anandasivam A, et al. Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers [J]. Journal of Parallel and Distributed Computing, 2011, 71(6): 732-749.
- [6] Mezmaz M, Melab N, Kessaci Y, et al. A parallel bi-objective hybrid metaheuristic for energy-aware scheduling for cloud computing systems[J]. Journal of Parallel and Distributed Computing, 2011, 71(11): 1497-1508.
- [7] Lucas-Simarro J L, Moreno-Vozmediano R, Montero R S, et al. Scheduling strategies for optimal service deployment across multiple clouds[J]. Future Generation Computer Systems, 2012(1):7.
- [8] Xu B M, Zhao C Y, Hu E Z, et al. Job scheduling algorithm based on Berger model in cloud environment [J]. Advances in Engineering Software, 2011, 42(7): 419-425.
- [9] Riakiotakis I, Ciorba F M, Papakonstantinou G, et al.

Towards the optimal synchronization granularity for dynamic scheduling of pipelined computations on heterogeneous computing systems [J]. Concurrency and Computation: Practice and Experience, 2012 (10): 2812.

作者简介:



吴煜祺 女, (1987—), 硕士研究生. 研究方向为并行分布处理、云计算.



曾国荪 男, (1964—), 博士, 教授, 博导. 研究方向为并行分布计算、可信网络软件.

(上接第 102 页)

作者简介:



王小花 女, (1987—), 硕士研究生, CCF 学生会会员 (E200018204G). 研究方向为无线 Mesh 网络.



李陶深 男, (1957—), 博士, 教授, CCF 高级会员 (E200050058S). 研究方向为无线 Mesh 网络、分布式数据库、网络路由算法.



葛志辉 男, (1978—), 博士, 副教授. 研究方向为无线网络.