

基于相空间的云计算仿真系统研究与设计

张磊¹ 王鹏¹ 黄焱^{2,3} 陈冬晓^{2,3} 郭又铭^{2,3}

(成都信息工程学院并行计算实验室 成都 610225)¹ (中国科学院成都计算机应用研究所 成都 610041)²
(中国科学院研究生院 北京 100049)³

摘要 提出并实现了一种基于热力学相空间的云计算集群仿真系统设计方法。将云计算集群节点信息投影到相空间,构建云计算系统的相空间分析模型,将云计算集群节点参数变化转变为其在相空间投影点的运动,并通过相空间的宏观热力学参数反映云计算集群的整体状态,为云计算核心调度算法的研究和云计算数据中心的运行测试提供了仿真测试平台和衡量标准。实验结果表明,全新构建的云计算仿真系统的仿真实验环境搭建简单,仿真结果能够直观、全面地反映出云计算集群的整体状态和调度算法性能等。

关键词 云计算,参数相空间,动量相空间,调度算法

中图分类号 TP391.9 **文献标识码** A

Research and Design of Cloud Computing Simulation System Based on Phase Space

ZHANG Lei¹ WANG Peng¹ HUANG Yan^{2,3} CHEN Dong-xiao^{2,3} GUO You-ming^{2,3}

(Parallel Computing Lab, Chengdu University of Information Technology, Chengdu 610225, China)¹

(Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610041, China)²

(Graduate University of Chinese Academy of Sciences, Beijing 100049, China)³

Abstract A design method for cloud computing simulation system based on thermodynamic phase space theory was proposed and realized. This method constructs the phase space analysis model of the cloud computing through projecting the information of the cloud computing cluster nodes onto the phase space, turning their parameter variations into the movement of the projected points in the phase space. Then the macroscopic thermodynamic parameter is introduced to reflect the general status of the cloud computing cluster. As per this method, a simulation test platform and a measurement standard are provided for researching on the cloud computing core scheduling algorithm and the operational test of the cloud computing data center. The test results show that cloud computing simulator system which is rebuilt completely has some main advantages as follows. It is easy to build experimental environment and simulation results are intuitive. This system can intuitively reflect the status of cloud computing cluster and the performance of the scheduling algorithm.

Keywords Cloud computing, Parameter phase space, Momentum phase space, Schedule algorithm

1 引言

云计算是建立在分布式计算、网络计算、分布式文件系统之上的一种效用计算模式^[1],为用户提供了海量的服务器资源。随着云计算产业的不断发展,针对云计算资源调度算法的研究也进行得如火如荼^[2]。文献[3]提出了云环境下基于伯格模型的作业调度算法。文献[4]提出了云计算环境下基于改进遗传算法的任务调度算法。目前为调度算法研究提供的云计算平台架构各异^[5],各有所长。云计算平台架构的不统一,直接导致了底层调度算法的研究不便于量化,不能形成统一的标准来衡量调度算法的优劣,同时追踪和分析复杂的

分布式云计算基础设施运行,也是一个非常具有挑战的问题^[6]。因此一个能够提供某一标准对调度算法进行衡量的云计算仿真平台就显得尤为重要。

目前,普遍使用的云计算仿真平台是由墨尔本大学开发的 Cloudsim 仿真系统^[7]。Cloudsim 是一种通用的、可扩展的模拟框架,支持大规模云计算基础设施,包括数据中心的仿真和建模。Cloudsim 最底层是基于 Simjava^[8]的离散事件仿真引擎;建立在 Simjava 离散事件仿真引擎上的是 Gridsim^[9],用于实现排队和事件处理、建立服务和主机等更上层的系统组件。Cloudsim 通过扩展 GridSim 层暴露的核心功能来支持以虚拟云为基础的数据中心环境。用户需要扩展

收稿日期:2012-05-15 返修日期:2012-08-23 本文受国家自然科学基金(60702075),广东省科技厅科技计划高新技术产业化科技攻关项目(2011B010200007),成都市科技局创新发展战略研究项目(11RKYB016ZF)资助。

张磊(1984-),男,硕士生,主要研究领域为云计算、并行计算、人工智能,E-mail:zhanglei2009bj@163.com;王鹏(1975-),男,博士,教授,CCF 高级会员,主要研究领域为云计算、并行计算、人工智能;黄焱(1982-),男,博士生,主要研究领域为云计算、并行计算、人工智能;陈冬晓(1986-),男,硕士生,主要研究领域为云计算、并行计算、人工智能;郭又铭(1986-),女,硕士生,主要研究领域为云计算、并行计算、人工智能。

Cloudsim 来实现更上层的功能。Cloudsim 虽然给出了云仿真的架构和实现,但主要存在以下两点问题。

(1) 仿真建模实验繁琐,用户不仅要编写自己的调度算法代码,而且要编写数据中心建模的代码。

(2) Cloudsim 虽然给出了基于云的建模方法和组件类,但并没有给出针对海量节点云数据中心的分析模型和云数据中心整体评估的评价指标。

不同于传统的集群系统,云计算集群是由海量服务器节点构成的,节点间因为频繁的计算迁移、计算备份、节点失效处理等高耦合性操作,所以具有了高耦合关系^[10],这与热力学系统由大量微观粒子组成,粒子间以任意方式相互作用的特性相似^[11]。根据前期研究成果即云计算集群与热力学系统具有相似性^[12]和 Cloudsim 仿真系统面临的一些问题,本文提出了基于相空间分析模型的云计算仿真系统的设计方法。该方法采用构建相空间的方式,将云计算集群的节点运行参数映射到相空间中,通过对云计算集群的相空间的广义热力学参数的分析来掌握集群的整体运行状况,进而对集群中当前运行的调度算法做出标准的衡量。从头构建和实现了人机交互界面、集群建模、负载发生器、任务调度算法库,用单台计算机来模拟海量云计算节点的运行,进而实现单一物理节点上做云计算调度算法的仿真实验。

构建云计算仿真系统的关键问题有以下几点:

- (1) 计算节点和计算任务数学模型的建立。
- (2) 调度算法管理平台。
- (3) 仿真结果分析模型的构建。

通过构建节点和任务的参数向量模型,提供调度算法接口和构建相空间分析模型的方法来解决上述问题。

本仿真系统具有以下几个特点:

(1) 仿真实验搭建过程简单,用户仅需根据自己的实验需求输入相对应的参数即可。

(2) 仿真结果动态显示在基于相空间的分析模型里,用户可以从整体上更加直观地了解调度算法的特点和适应的场景。

最后本文通过对比不同调度算法在相空间的实验结果,体现了该仿真系统对云计算集群性能的宏观掌控能力。

2 相空间分析模型的构建

在热力学中,相空间是一个用来表示系统所有可能状态的空间,系统每个可能的状态都有对应的相空间的点^[13]。

隔离地观察集群中单个服务器节点并不能反映出云计算集群的整体情况,从而不能对当前运行的调度算法进行衡量。为了从宏观上观察当前集群的这个整体情况,以集群中节点的 n 个状态参数如 CPU 占用率、内存占用率等为坐标轴,构成一个 n 维空间;将集群中的节点的参数信息状态映射到此空间中,每个节点表现为一个点,点在 n 维空间中的位置反映了节点当前的运行情况,这样就构成了云计算集群的相空间,将服务器节点的状态变化转变为相空间中投影点的运动,对云计算集群的整体研究也就随之转变为了对相空间中广义热力学参数的研究。

2.1 参数相空间

调度算法在资源分配方面的能力能够通过云计算集群的整体负载均衡性反映出来。在良好的调度策略下,节点负载

差异不应过大,整个集群负载处于均衡状态。为了直观地反映出集群的整体负载情况,本文构建了云计算系统的参数相空间。将服务器的两个或多个状态参数定义为广义坐标轴,构建成的二维或多维空间则为云计算系统的参数相空间。再将当前集群中的节点根据自身负载情况投影到参数相空间中,则对整个集群的负载均衡性研究就转变为了对参数相空间中投影点分布情况的研究。本文以 CPU 负载和内存占用率为例,以内存占用率作为 x 坐标轴,CPU 负载作为 y 坐标轴,构建参数相空间,如图 1 所示。

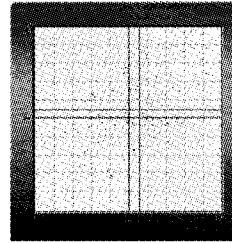


图 1 仿真系统参数相空间的构建

集群在参数相空间中投影的重心位置是判断集群平均负载的重要参数。以二维相空间为例,定义重心坐标 $G(x_0, y_0)$

的计算公式为: $x_0 = \frac{\sum_{i=1}^m x_i}{m}$, $y_0 = \frac{\sum_{i=1}^m y_i}{m}$, m 为云计算集群的总节点数, (x_i, y_i) 为第 i 个节点投影在相空间中的坐标。即重心坐标等于集群节点的平均坐标,图 1 中心位置上方的十字交叉点即为当前集群的重心位置。

2.2 动量相空间

除了通过相空间的重心来衡量系统的平均负载情况,整个云计算集群的运行稳定状况是衡量调度算法的另一个标准。为了反映出集群的运行稳定性,在参数相空间的基础上,再以节点在参数相空间中投影的位置离重心的距离为 x 坐标轴,节点的动量为 y 坐标轴,构建云计算集群的动量相空间。

假设在 $t-1$ 时刻某一节点投影在参数相空间内的坐标为 (x_{t-1}, y_{t-1}) , t 时刻投影坐标为 (x_t, y_t) , 则 t 时刻该节点的运动速度为 $v_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$ 。定义在由同质服务器构成的云计算集群中,节点动量的计算方法为: $p_i = \frac{v_i}{\sqrt{2}}$ 。其中, v_i 为第 i 个服务器单位时间内在参数相空间中的移动距离。构成的动量相空间如图 2 所示。

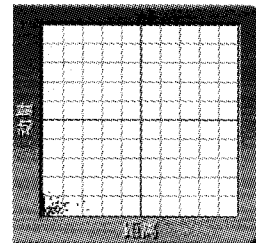


图 2 仿真系统动量相空间的构建

在动量相空间中投影点越靠近原点,表明系统当前的动态工作状态越平稳,负载均衡状态也越好。

2.3 广义宏观热力学参数定义

参数相空间与动量相空间直观反映了当前节点的状态,为了验证相空间对调度算法结果显示的正确性,本节引入热

力学中具有明确物理意义的宏观量来对云计算集群的工作状态进行描述和分析。

2.3.1 云计算系统的广义温度

在热力学系统中,温度是一个重要的热力学参数,与分子运动的平均速度有关,温度越高的系统分子运动的速度越快。

在云计算集群的二维相空间中,假设任一时刻外部负载请求数为 1,对集群请求参数输入为 a_i, b_i ,同时刻集群释放的任务数为 n ,释放任务数的负载为 c_j, d_j ,则拥有 m 个节点的集群总平均速度 \bar{V} 的大小为:当前间隔内服务器所有投影点由于释放任务造成的平均运动速度+外部请求造成的节点的平均运动速度,即:

$$\bar{V} = \frac{\sum_{i=1}^n \sqrt{c_j^2 + d_j^2}}{m} + \frac{\sum_{i=1}^l \sqrt{a_i^2 + b_i^2}}{m}$$

定义集群广义温度 $T = \bar{V}$,则集群外部请求给节点带来的具有耦合性的运动和节点内部的运动共同决定了温度的大小,节点运动越激烈,温度越高。拥有良好调度策略的集群应处于稳定的运行状态,节点运动速度比较平缓,集群温度 T 应较低。因此,集群温度的高低充分反映了当前集群运行的稳定状况。

2.3.2 云计算系统的广义熵

在热力学中,熵反映了热力学系统的混乱程度。为了从理论上验证相空间对集群负载反映的正确性,定义集群的广义熵为: $K = \frac{\ln l}{\ln(n \times n)}$,其中 $n \times n$ 表示二维参数相空间横纵坐标各划为 n 等分,使相空间为 $n \times n$ 的网格; l 为云计算服务器投影在参数相空间的网格数。

容易得知 $1 \leq l \leq n \times n$,所以广义熵的取值范围为 $[0, 1]$ 。当服务器全部落入同一个网格内,即 $l=1$ 时,集群负载处于理想的均衡状态,此时广义熵值为 0;当服务器散布在所有的网格内,即 $l=n \times n$ 时,集群负载处于完全不平衡的状态,此时广义熵值为 1。可以看出,集群的广义熵值与相同任务请求下调度算法的优劣成反比关系,即熵越大调度算法越差,反之亦然。广义熵的定义从理论上对运行的调度算法进行了衡量,也为相空间显示的正确性提供了理论依据。

3 云计算相空间仿真系统设计

仿真系统主要包括 4 个部分:基于相空间的仿真结果分析及人机交互界面、负载发生器、任务调度算法库、集群建模。用户通过人机交互界面设定数据中心规模参数、任务负载压力等仿真实验参数。在算法库里选择两种不同的算法即可开始对比实验。算法可根据需要进行动态添加。整体架构如图 3 所示。

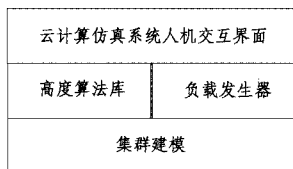


图 3 云计算相空间仿真系统架构图

3.1 集群建模

集群建模是实验生命周期内在单台主机上仿真云数据中心的一组主机。主机代表云计算集群中的一个物理节点,用一组参数向量来表示它的工作状态(如 CPU 占用率、内存占

用率、网络占用率...)。假设有 n 个任务在当前节点运行,用 T_i 表示此节点第 i 个任务的 CPU 占用率,则:

$$\text{CPU 占用率} = T_1 + T_2 + \dots + T_n$$

其他分量具有类似的计算公式。

因为占用率是个相对值,所以各个分量的取值范围为 $[0, 1]$ 。如果将要分配的任务的资源占用率与当前节点的资源已使用率之和超过 1,则由调度算法将任务分配到其他节点。采用相对值来描述节点的状态,避免了计算机性能的差异,对算法性能的对比量化更具理论意义。

把节点向量映射到相空间中,通过相空间投影点的变化动态反映出集群当前节点的状态变化。在本仿真系统中,选择向量中的任意两个分量映射到二维相空间中,这样对云计算系统的研究就转变为了对二维相空间点的研究。

3.2 负载发生器

负载发生器是整个仿真系统的关键部分,其作用是产生虚拟任务。虚拟任务同样是用一组参数向量来表示,与节点参数向量不同的是虚拟任务向量多一个时间分量,其代表当前任务在当前资源占用率的情况下需要运行多长时间才能完成。时间分量也是相对时间,仿真实验运行的过程中,每经过一个时间片,所有任务的时间占用减 1。每个最小时间间隔用 T_0 来表示。当任务的时间片减为 0 时,表明任务执行完毕,就释放掉这个任务,以备接受新的任务分配。

不同类型的任务对节点资源的占用差异较大,反映在调度算法上的均衡性就会有较大差异。为了仿真不同的任务模型,采取限定虚拟任务向量分量上下限的策略。表 1 展示了在 100 个节点的集群规模下,不同任务模型的参数设定。

表 1 任务模型参数设定

任务模型	节点数	请求数/ T_0	占用率	时间片
低负载 高并发	100	2000	0~0.01	1~10
常规负载 常规压力	100	200	0~0.1	1~10
高负载 低并发	100	16	0~1	1~10

参照表 1 的参数设定,就可以针对实验需要的不同设定虚拟任务参数。

真实系统任务的到达时间具有一定的随机性。均匀分布与高斯分布为排队论中最常见的两种输入过程的分布,因此本文为了模拟真实系统的随机性,提供了这两种分布模型供选择。负载发生器根据用户指定的仿真策略产生虚拟任务,并将虚拟任务存入任务队列等待调度算法的调度。

3.3 任务调度算法库

任务调度算法库里实现了几种常见的调度算法来作为仿真实验的对比算法,同时用户自己实现的调度算法也要编译进这个算法库里以备仿真实验时调用。实现的常用算法包括轮询调度算法、最小负载优先调度算法。

本文中用户算法的实现接口函数为:

SchedulingAlgorithm(CSSCloudTask *pCSSTask, CSS-CloudCluster *pCSSCluster)

其中,pCSSTask 指向任务队列,pCSSCluster 指向虚拟集群。

调度算法不断地从任务队列取出任务,并按照一定的策略分配给集群的某个节点。

(下转第 111 页)

- [3] 陈阳舟, 田秋芳, 张利国. 基于神经网络的城市快速路交通拥堵判别算法[J]. 计算机测量与控制, 2011, 19(1): 167-169
- [4] 刘建美, 马寿峰, 贺正冰, 等. 控制与诱导的协调中路网拥堵状态识别方法[J]. 管理科学, 2010, 13(11): 35-40
- [5] 林庆, 张永鑫, 吴旻. 智能交通系统综合信息平台的设计[J]. 计算机工程与设计, 2008, 29(13): 43-44
- [6] Gao Zi-you, Zhao Xiao-mei, Huang Hai-jun, et al. Research on

- [7] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks [J]. Nature(S0028-0836), 2007, 407(6804): 651-654
- [8] 郎益顺. 转换 OD 的动态交通分配模型研究[D]. 哈尔滨: 哈尔滨工业大学, 2009: 25-29
- [9] 熊丽音, 陆锋, 陈传彬. 城市多模式交通网络特征连通关系表达模型[J]. 武汉大学学报: 信息科学版, 2008, 33(4): 393-396

(上接第 86 页)

4 测试与分析

为了验证该系统在云计算调度算法衡量中的有效性和可行性, 本文采用最小负载优先调度算法和轮询调度算法来进行测试。将最小负载优先调度算法和轮询调度算法应用到仿真系统中, 对比两种算法在相同的负载压力下对集群负载均衡的影响。具体实验参数设定如表 2 所列。

表 2 仿真实验参数设定

节点数	调度算法	请求数/ T_0	占用率	时间片
100	最小负载优先	200	0~0.1	1~10
100	轮询算法	200	0~0.1	1~10

其中, 占用率与时间片的产生均采用符合均匀分布的负载发生策略。当仿真实验达到稳定状态时, 仿真实验结果如图 4 所示。左侧为采用最小负载优先调度算法的负载情况, 右侧为采用轮询调度算法的结果, 中间分别显示了通过计算获得的在两种调度算法下集群的广义温度和广义熵值。

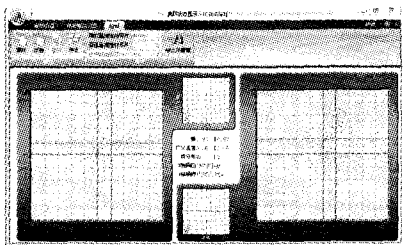


图 4 仿真实验用户界面截图

任务请求数为 $200/T_0$, CPU 和内存占用率均为 $0 \sim 0.1$, 仿真了常规负载常规并发的任务模型。

从图 4 可以看出, 在当前情况下, 最小负载优先调度算法(左侧)下的节点分布更为集中, 在负载均衡性方面更具有优势。此时最小负载优先调度算法的熵值为 0.73, 也较轮询调度算法的 0.78 低, 这从理论上验证了相空间显示的正确性。

动量相空间图反映了在当前调度算法下集群的运行稳定性。由图 4 两幅上下对应的小图可以看出, 最小负载优先算法与轮询算法的广义温度均为 0.08, 表明最小负载优先调度算法和轮询调度算法在动量图上几乎没有差别, 集群的运行均较为稳定。

通过上述实验, 可以看出该仿真系统的特点如下:

1. 利用相空间分析模型给出了衡量调度算法的统一标准, 使其具有很强的分析能力和表现能力。
2. 支持对数据中心及不同类型任务的仿真建模。

3. 对比实验效果显著, 为调度算法的研究提供了很好的仿真平台。

该仿真系统和分析方法, 使系统调度算法的研究人员可以专注于对调度算法的研究, 而不用关心云计算的底层细节。

结束语 本文介绍了一种基于相空间的云计算仿真系统架构模型。实验表明, 该系统为云计算调度算法的研究提供了很好的验证测试平台, 并为调度算法的衡量提供了统一标准。下一步将对仿真系统进行扩展, 将相空间分析模型扩展到多维空间。

参考文献

- [1] Armbrust M, Fox A, Griffith R, et al. Above the Clouds; A Berkeley View of Cloud Computing [R]. Tech Rep; UCB/EECS-2009-28. University of California at Berkeley, 2009
- [2] 周海花, 华薇娜. 国内外云计算研究的现状与发展[J]. 现代情报, 2012, 32(2): 26-35
- [3] 赵春燕. 云环境下作业调度算法研究与实现[D]. 北京: 北京交通大学, 2011
- [4] 李建锋. 云计算环境下基于改进遗传算法的任务调度算法[J]. 计算应用, 2011, 31(1): 184-186
- [5] 陈康, 郑伟民. 云计算: 系统实例与研究现状[J]. Journal of Software, 2009, 20(5): 1337-1348
- [6] 李乔, 郑啸. 云计算研究现状综述[J]. 计算机科学, 2011, 38(4): 32-37
- [7] Calheiros R N, Ranjan R, et al. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms[J]. Software—Practice and Experience, 2011, 41(1): 23-50
- [8] Howell F, Mcnab R. SimJava: A discrete event simulation library for java[C]//Proceedings of the First International Conference on Web-based Modeling and Simulation. 1998
- [9] Buyya R, Murshed M. GridSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing[J]. The Journal of Concurrency and Computation: Practice and Experience (CCPE), 2002, 14: 13-15
- [10] 王鹏. 走近云计算[M]. 北京: 人民邮电出版社, 2009
- [11] 王竹溪. 热力学(第 2 版)[M]. 北京: 北京大学出版社, 2005
- [12] 王鹏. 云计算相空间广义热力学参数定义及分析[J]. 计算机应用, 2012, 32(8): 2172-2175
- [13] 王克, 范猛. 泛函微分方程的相空间理论与应用[M]. 北京: 科学出版社, 2009