

# 大数据下虚拟化与云计算的应用分析

周家栋

(江苏省盐城技师学院, 江苏 盐城 224000)

摘要: 大数据已经为人们的生活带来了许多的改变, 云计算的热度可以说是蒸蒸日上, 因为云计算是信息技术领域中非常重要的技术, 所以云计算可以不断地推动信息产业的持续发展, 并且能够促进社会经济的全面快速发展。而虚拟机技术是云计算的重要基础, 其为云计算的发展提供了很大的帮助。笔者在这篇文章中详细地介绍虚拟机技术以及云计算的研究方向和具体应用。

关键词: 大数据; 虚拟机技术; 云计算; 应用; 探究

中图分类号: TP393 文献标识码: A

文章编号: 1009-3044(2021)08-0178-02

DOI: 10.14004/j.cnki.ckt.2021.0682



开放科学(资源服务)标识码(OSID):

虚拟机技术对于云计算的发展是至关重要的, 是帮助云计算实现自身价值的重要技术, 比如, IaaS 架构, 是云计算资源池的核心构想。虚拟机技术的作用是对云计算的数据进行有效的管理, 特别是要管理好关于负载均衡的资源。同时, 虚拟机的动态迁移技术又占据着虚拟机技术的重要地位。

## 1 虚拟机动态迁移技术

虚拟机的动态迁移技术可以提高云计算的效率, 同时, 其能够与多任务部署方法进行有机的结合, 这样便可以更好地分析处理大数据。笔者在此提出一种节能感知的动态迁移技术, 并提供这种技术的核心理念、方法过程以及测验分析。

## 2 节能感知虚拟机的动态迁移的位置选择策略 PS-ABC

### 2.1 系统的结构方案

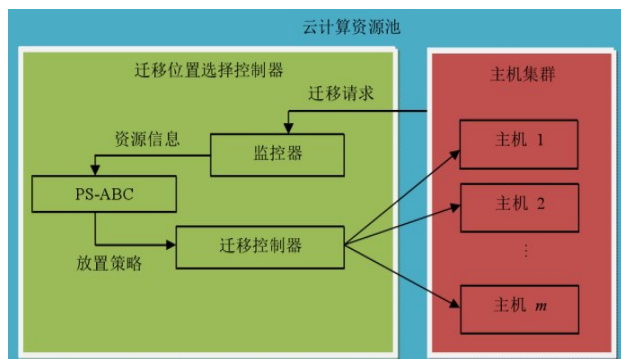


图1 PS-ABC 系统架构

如图1所示, 其体现出云计算当中的虚拟机动态迁移系统的结构, 其具体说明了用于位置迁移的控制器PS-ABC的位置以及实体之间的复杂关系。在一个时间段中, 虚拟机的动态迁移的信号请示会传递到监控器中, 并且虚拟机的动态迁移请示

会根据可用计算资源的变化而变化。当一个时间段结束的时候, 监控器把所接收到的请示传递到控制器当中。接下来, 控制器通过研究分析所得出的方案以及有效的信息合成放置的方法。再然后, 控制器会向迁移控制器提供放置的方法, 控制器便会控制虚拟机的动态迁移。最后, 虚拟机会被迁移到目标主机上。

### 2.2 解的表示

为了得到一种方法, 其可以找到一个时间段中收集的等待迁移的虚拟机的目标主机, 而这种方法是依靠人工蜂群算法的。首先面临的问题就是如何表示解, 它直接反映了问题域以及人工蜂群算法中的食物源之间的关系。在使用依赖人工蜂群算法的方法的过程中, 食物源代表了解决一个特定问题的具体办法。在整个过程中,  $m$  个物理主机要接受  $n$  个虚拟机, 所以, 所提出的问题是  $n$  维的问题, 也就要求食物源被划分为  $n$  维解向量。同时, 每一个独立的维度都存在一个小于等于  $m$  的正整数集合, 而这个集合就成了可供选择的位置编号, 而向量解就等同于食物源, 此外, 多维向量在数学中有着特定的表示方法, 所以可以用数学符号表示等待迁移的虚拟机的目标主机的编号。

### 2.3 PS-ABC 的工作理念

PS-ABC 方法是一种虚拟机动态迁移的一种策略, 其比较新颖且启发性高, 可以选择虚拟机动态迁移的目标位置。PS-ABC 方法是在改进版的人工蜂群的基础上, 提出了均匀随机初始化的理念, 以便能够改善食物来源的初始化进程, 最终 PS-ABC 方法便具备搜索全部内容的能力。除此之外, PS-ABC 方法利用二分搜索理念, 以便减少雇佣蜂产生候选食物位置所需要的时间, 最终使搜索效率变得更高。与此同时, PS-ABC 方法充分利用玻尔兹曼选择策略的优势来帮助跟随蜂实现概率选择, 这样能够使 PS-ABC 方法具备自我适应的能力, 以便适时地调整压力, 这样可以提高探索全局的能力以及后期的局部收

敛能力。最后,笔者使PS-ABC方法吸收贝叶斯理论的精华之处,这样就能够提高系统的准确度。

### 3 PS-ABC 的实验与分析

通过实验研究提出了关于PS-ABC方法的性质以及工作效率,本文通过Cloud Sim toolkit 事件驱动模拟器模拟了动态的云计算平台。Cloud Sim框架可以计算出系统所消耗的能源,而它的计算就是在Cloud Sim平台是进行的。利用这个平台,本文把PS-ABC方法、随机迁移策略、基于遗传算法的最优化迁移策略这三者放在能耗以及无效虚拟机迁移次数方面进行了详细的比较分析。本文从以下三个方面测验了PS-ABC的性能,并且有实验结果得出PS-ABC能够提升云计算平台的性能。

#### 3.1 实验背景

通过模拟云平台建立一个动态云资源池,其中含有了大量的各种各样的物理主机。并且通过Cloud Sim平台安排了几十轮虚拟机动态迁移的请求工作,同时,每一轮都包含了15个主机不同且资源不同的迁移请求。

#### 3.2 PS-ABC、随机迁移策略以及基于遗传算法的最优化迁移策略在所消耗的电能的数量方面的比较

实验分别布置了PS-ABC、随机迁移策略以及基于遗传算法的最优化迁移策略的模拟云计算平台,并且通过它们各自所消耗的能量来证明PS-ABC方法是十分节能的。在本实验过程中,设定一个时间段为400秒,一小时当中均匀分布着各个批次的迁移事件,其中包含了9个时间段。除此之外,主机资源每隔半个小时都会变化一次。

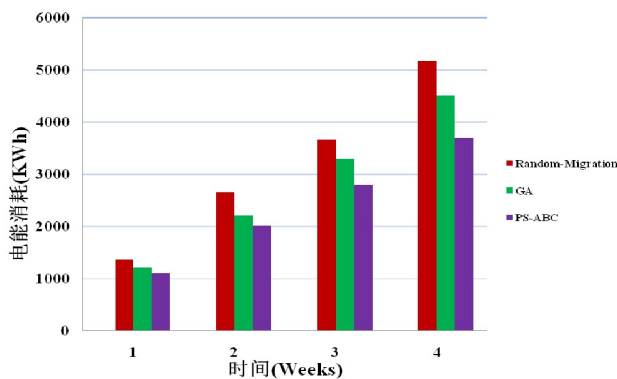


图2 三者电能消耗对比

部署PS-ABC的云计算平台在一周当中所消耗的电能最少,除此之外,在两周的时间内,部署PS-ABC的云计算平台所消耗的电能增量是最少的,但是,随机迁移策略在两周内的电能增量有着明显的增加。而出现这种情况的原因就是随机迁移策略对云环境的变化适应能力不强,于是便有可能选择不合适的目标主机,所以就会增大能源消耗。而且,随着时间的逐渐推移,能源消耗的差别也会越来越大。尽管遗传算法被人们广泛运用,但是,相较于PS-ABC,遗传算法的编码以及解码都要根据不同的个体做出相应的改变,也就是说,遗传算法仅仅适用于处理规模比较小的问题。根据算法以及处理过程可以发现,PS-ABC的性能比遗传算法要好,所以,PS-ABC能够更好地节省电能。

#### 3.3 虚拟机迁移事件的失败率

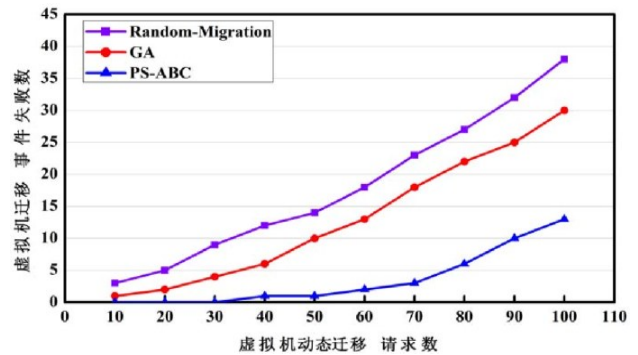


图3 迁移失败率对比

实验通过Cloud Sim调度主机故障以及主机关闭来进行动态主机失去作用的模拟,由于被选择的主机无法对虚拟机的迁移请求做出回应,所以上述事件会导致虚拟机迁移的失败。这次试验是比较PS-ABC、随机迁移策略以及遗传算法的虚拟机迁移失败数量与虚拟机动态迁移请求数量之间的关系。根据对比可以看出,部署随机迁移策略以及遗传算法的云计算平台的虚拟机动态迁移的失败的数量会随着动态迁移请求的增加而增加,因为在整个过程当中,很多内存数据是过时的,所以云计算平台不能根据环境的变化而做出适当的改变。而PS-ABC方法中融合了玻尔兹曼选择策略和贝叶斯定理,所以其做出反应所需要的时间比较短,对动态云的适应能力很强,并且PS-ABC可在极短的时间内找到可供使用的主机,并且该主机能够提供动态迁移所需要的资源。正如上文所阐述的,PS-ABC可以帮助虚拟机的动态迁移,并且,在整个过程中所消耗的能量都是非常少的。

#### 3.4 遗传算法(GA)与PS-ABC各自的加速效果

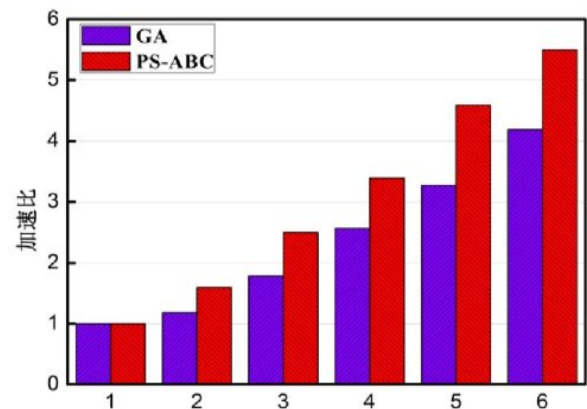


图4 加速比对比

在这个实验当中,可以根据增加的处理器的数目来估算GA与PS-ABC的加速比,且加速比的公式可以表示为:

$$S_q = \frac{T_1}{T_p}$$

公式当中的 $T_1$ 指的是算法在处理器条件下的执行时间, $T_p$ 指的是在数目为 $p$ 的处理器是处理工作负载所需要的时间。所以,加速比 $S_q$ 可以表示程序进行并行的性能以及效应。根据对比,PS-ABC以及GA的加速比都会随着处理器数目的增加而

(下转第184页)

由表 2 可知,KDPCK-means 算法得到的初始类心、最终类心与 Iris 数据集真实类心十分接近。

表 3 各个算法的精确度(单位:%)及迭代次数的对比

| 数据   | K-means |      | 最大最小距离聚类算法 |      | KDPCK-means 算法 |      |
|------|---------|------|------------|------|----------------|------|
|      | 精确度     | 迭代次数 | 精确度        | 迭代次数 | 精确度            | 迭代次数 |
| Iris | 54.80   | 6    | 88.93      | 8    | 96.67          | 2    |
| Wine | 50.34   | 7    | 68.65      | 11   | 97.19          | 3    |

在这两种数据集中,KDPCK-means 算法的精确度显优于前两种算法,其迭代次数也明显少于前两种算法。通过以上实验数据可知,KDPCK-means 算法在一举解决了原始 K-means 算法中问题的同时,在准确率、稳定性及运行效率上都得到了有效的提升。

参考文献:

[1] Al Hasib A,Cebrian J M,Natvig L.A vectorized k-means algorithm for compressed datasets:design and experimental analysis [J].The Journal of Supercomputing,2018,74(6):2705-2728.

[2] Douzas G,Bacao F,Last F.Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE[J].Information Sciences,2018,465:1-20.

[3] García J,Crawford B,Soto R,et al. A k-means binarization framework applied to multidimensional knapsack problem[J]. Applied Intelligence,2018,48(2):357-380.

[4] Manju V N,Lenin Fred A.AC coefficient and K-means cuckoo optimisation algorithm-based segmentation and compression of compound images[J]. IET Image Processing, 2018, 12(2): 218-225.

[5] 陈思敏. 基于位置指纹识别的 WiFi 室内定位算法研究与实现[D]. 南京:南京邮电大学,2016.

[6] 韩雅雯.kmeans 聚类算法的改进及其在信息检索系统中的应用[D]. 昆明:云南大学,2016.

[7] 孔超. 基于分布式平台的高校网络舆情分析系统研究与实现[D]. 成都:电子科技大学,2017.

[8] 郭靖.K-means 聚类算法改进研究[D]. 北京:中国人民公安大学,2018.

【通联编辑:李雅琪】

(上接第 179 页)

增加,并且 GA 的加速比在任何时刻都是小于 PS-ABC 的加速比的,并且 PS-ABC 的加速比的图像近似符合线性函数的图像。对于 PS-ABC 而言,在逻辑的层面上,其可以把整个搜索空间划分为多个独立的子空间。同时,这些子空间分别对应着自己的搜索程序,并且执行者自己的搜索工作,接着,雇佣蜂共享大量的信息资源。GA 的机制比较适合集中处理,而 PS-ABC 可以进行并行处理,从这些特点就可以看出来,PS-ABC 的测试性能要比 GA 的性能好得多。此外,PS-ABC 的工作性能将会在处理大数据的时候充分地体现出来。

4 结语

PS-ABC 可以减少系统的能源损耗以及提高服务性能的稳

定程度,除此之外,PS-ABC 还能够确保虚拟机动态迁移的性能良好。最终,各项结果都证明了 PS-ABC 有能力配合虚拟机的动态迁移,从而提高大数据的分析处理水平。

参考文献:

[1] 姚金玲,李莉. 浅谈大数据背景下虚拟化与云计算的优势及应用[J]. 电脑知识与技术,2020,16(21):44-45.

[2] 江璜. 基于 VMware vSphere 的虚拟化资源池应用研究[J]. 软件工程,2020,23(3):32-34.

[3] 甘云志. 虚拟化技术在新一代云计算数据中心的应用研究[J]. 数字技术与应用,2020,38(1):70,72.

【通联编辑:先文玲】