

云存储环境下副本选择策略研究

张翠苹 郭振洲 拱长青

(沈阳航空航天大学计算机学院 沈阳 110136)

摘要 云存储服务提供商为了满足各类云用户的存储需求,一般采用划分固定大小的数据块、冗余备份等技术来存储数据,关于块放置、最佳副本选择、副本粒度等存储机制的研究一直是加快大文件存取速度的重要内容。面向云存储系统中存储节点的异构性,设计了一种采用层次分析法对节点性能指标加权并依据加权指标改进粒子群算法的策略(AHPPSO)。通过引入与存储节点性能相关的加权评价矩阵,使得粒子群算法向综合性能较高的节点进化,在不增加存储空间成本的基础上,加快了存取数据的速度。在自主搭建的云存储系统中实现了该策略,实验结果显示该策略能够适应多种用户需求,并且在一定程度上实现系统负载均衡。

关键词 云存储,异构性,粒子群算法,层次分析法,负载均衡

中图分类号 TP302 **文献标识码** A

Study on Strategy of Replica Selection in Cloud Storage Environment

ZHANG Cui-ping GUO Zhen-zhou GONG Chang-qing

(School of Computer, Shenyang Aerospace University, Shenyang 110136, China)

Abstract In order to meet the needs of various users for cloud storage, cloud storage service providers generally divide the data into fixed size block and use redundancy backup technology to store data. So the researches on storage mechanism of the block placement, the best replica selection and the replica size have always been hot spots in improving the transmission speed of big file. According to the heterogeneity of storage nodes in cloud storage system, the improved strategy which uses AHP to weight the index of node performance, and uses the weighted index to improve particle swarm optimization algorithm(AHPPSO) was proposed. By introducing the weighted evaluation matrix associated with the performance of the storage node, PSO evolves toward the node of high comprehensive performance, improving the data transmission speed without increasing the cost of storage space. The strategy was realized in a self-built cloud storage system, and the experiment result shows that this strategy can adapt to the various needs of users, and achieve the system load balancing to a certain extent.

Keywords Cloud storage, Heterogeneity, Particle swarm algorithm, AHP, Load balancing

1 引言

随着云计算规模的不断扩张,在逐步提高云计算的计算能力和传输速度的同时,催生了企业或个人对“存储即服务”模式的需求,给云计算系统的资源存储管理带来了更为严峻的挑战。从服务层面上来讲,“存储即服务”是通过把服务资源部署在云计算的分布式存储系统中,方便用户可以随时随地得到所需存储服务的一种重要手段,目前这种模式逐渐成为从云计算领域衍生出来的另一个研究热点方向——云存储。云存储^[1]指的是通过用户终端应用程序将网络中多种类型的存储设备连接起来并交互信息实现协同工作,以此实现共同为云用户或者云应用提供数据存储和访问等底层存储服务的一种模式。与云计算平台不同的是,为了提高数据存储服务能力,云存储侧重关注通过优化资源分配、整合和调度策略来增强数据存储服务能力。由于云用户的需求不同,通常

来说用户仅关注整体文件系统中自己感兴趣的小部分内容。基于这种实际情况,对整个系统采用完全备份机制不但浪费大量存储空间,而且会很大程度上影响系统的存储和访问效率。合理高效的资源管理策略是提高系统存储服务水平的重要方法。服务成本代价和服务性能成为目前云存储服务领域中两个突出的研究方向,并且在商业领域和学术领域都已经取得了非常不错的成果。

目前得到广泛应用的分布式存储系统包括用于商业应用的 Google File System(GFS)、Amazon Simple Storage Service(S3)和用于科研的 Hadoop Distributed File System(HDFS)^[2]。为了保障数据的可用性和可靠性^[3],Google、Amazon 和 Hadoop 都默认采用随机“三副本”的数据备份冗余机制。通过在不同地理位置建立自身的副本,存储系统可以利用数据的冗余性和地理分散性提高系统的容错能力和数据的可靠性。

本文受辽宁省教育厅科学基金(L2013064),中航工业技术创新基金(基础研究类)(2013S60109R)资助。

张翠苹(1990—),女,硕士生,主要研究方向为分布与并行计算,E-mail:beauty1001@yeah.net;郭振洲(1977—),男,博士,讲师,主要研究方向为密码学与信息安全;拱长青(1965—),男,博士,教授,主要研究方向为云计算。

此外,在学术领域针对降低云存储服务存储空间成本的研究中,文献[4]提出一种控制最小副本数量的策略,旨在节约系统存储资源。但是这种策略需要人为设定数据可用性,在规模庞大的云存储环境下,数据可用性是多变的,因此付出的代价是巨大的。从存储代价的角度出发,文献[5]提出主动式检查副本的可用性和维护系统的可靠性,仿真实验表明,较“三副本”策略,这种策略有效地提升了存储空间利用率。但其需要系统人为设定文件的可用性,这在节点数量庞大的云系统中不是一个理想的解决方案。文献[6]中将服务质量感知数据副本问题转化为最小代价最大流问题,并且提出了节点连接技术来减短数据复制时间。当前一些分布式存储平台,如 Luster, PVFS, GFS 等,都采用纠删码技术提高系统的容错能力,并且在文献[7]中针对 HDFS 文件系统数据块的特征,改进了纠删码编解码算法,提高了算法的执行效率。文献[8]中提出了一种优化数据冗余策略 RIRS(Replication Improved RS),该策略将完全备份和改进的 RS(Read-Solomon)纠删码两种数据冗余技术结合起来,兼备了完全备份机制的低时延和纠删码技术高可靠性的优点,并可以节省大量存储空间。文献[9]中指出在分布式存储系统中,存储介质数目多,纠删码的重构效率是影响系统性能的关键,而存储利用率并不是关注的焦点。

本文的主要工作包括以下两点:

(1)针对云存储环境下的资源分配和调度问题,对现有的一些解决方法进行分析总结。针对副本选择问题构建了数学模型,采用层次分析法对存储节点的评价指标进行了从定性到定量的转化处理。

(2)针对云用户的需求存在差异化的问题,设计了基于用户需求相关的云存储副本选择方案。该方法使云提供商可以依据云用户的不同需求提供差异化、个性化的存储方案。

2 相关工作

随着云存储服务模式的不断推进,云平台的服务性能和服务代价问题已经成为了云提供商更为关注的问题。目前,针对云环境的资源管理问题主要有 3 类解决方法,分别是基于预测的调度策略^[10,11]、基于完全备份和纠删码相结合的备份策略和基于进化算法的调度策略。

(1)基于预测的备份策略

根据云存储环境下用户访问特征,文献[10]中根据“热点数据聚集”现象提出最佳机架策略。但是模型中未考虑对模型性能影响较大的一个因素,即副本的放置位置。

根据云存储环境下数据节点的异构性,文献[11]中依据相似环境下副本具有相似的服务能力水平,提出一种基于环境相似程度和用户服务质量满意度的多约束条件下的副本选择策略。但是该策略没有对未评价过的副本和已被评价过的副本进行综合考虑。

(2)结合完全备份和纠删码的备份策略

基于纠删码编解码算法的备份策略具有算法易于实现、容错能力强和存储资源消耗低等优势^[12,13]。如文献[12]提出一种块副本和基于纠删码相结合的存储节点容错技术。对访问频率较高的热点数据进行块副本存储,对非热点数据进行基于 RS 编解码的冗余存储。但是这种编码机制有一定的缺陷,它的编解码速度较慢,导致备份机制的性能下降。在云存储系统中,存储介质可被认为是无限扩展的,纠删码的编解

码效率是影响系统性能的关键,而存储资源的利用率并不是关注的重点。

(3)基于进化算法的调度策略

文献[14]副本选择过程中采用的是进化算法中的蚁群算法,副本和存储节点的信息特征值直接影响最优路径的选择,而种群历史最优位置的选择情况将会影响下次迭代过程中个体的信息素值。但是蚁群算法有其自身的缺点和局限性,如果不对其改进,在副本选择问题求解上易陷入局部最优。

文献[15]中提出了一种基于改进遗传算法的副本选择与创建的策略。但是该策略有一定的弊端,创建过多的副本而不删除访问热度降低的副本不仅浪费大量存储资源,而且会影响系统的性能。

综上所述,基于预测的备份策略具有预测准确度高、计算速度快的特点,可以满足云平台即时计算、即时存储的要求,但是需要大量的先验知识。基于完全备份和纠删码相结合的备份策略具有较高的容错性,同时相比于多副本机制来说,其可以大大减少备份对系统资源的消耗。但是在云存储系统中,存储介质被认为是可以无限扩展的,存储资源的利用率并不是所有云用户最关心的问题,而纠删码的编解码效率成为影响系统性能的关键。基于进化算法的备份策略的优势在于此类算法运行速度快,能够满足云存储系统对数据的高实时性的要求。进化算法一般在全局优化问题求最优解中能够获得一个误差很小的近似最优解,因此可以满足云环境下实际应用的需求。因此本文针对云存储系统的负载均衡问题,设计了一种基于粒子群算法的副本选择策略,并对此展开研究工作。

3 问题模型及数学描述

本节首先介绍了云存储系统模型,然后对云存储系统中的副本问题的相关概念进行阐述,并给出定性目标到定量转化的处理方法。

3.1 系统模型

分布式数据存储模型如图 1 所示。

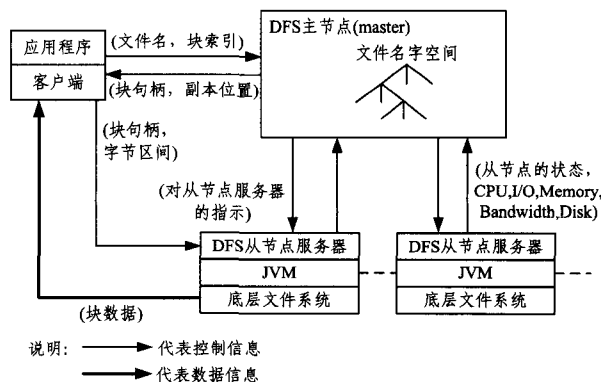


图 1 DFS 系统模型

其中,客户端是使用云存储平台完成相关工作的用户;DFS 主节点负责接收并处理用户任务请求,以及检测各个从节点的运行状态;DFS 从节点服务器表示存储数据的节点,负责实时获取硬件各项配置及性能并以“心跳”方式发送给主节点。在副本放置策略中,主节点收到用户请求后,会根据从节点的实时系统状态合理地选择数据副本的存放位置。在副本选择策略中,主节点根据从节点发送过来的硬件性能和网络状态等信息合理地为用户选择一个最佳副本。

DFS从节点发送给主节点的信息包括系统当前的CPU、内存、磁盘、带宽等状态信息,间隔一定时间获取信息并间隔一定时间发送给主机点,保证了主节点端信息的实时更新。

主机点服务器检测从节点的运行状态,实时更新节点列表,当扩展新节点或节点失效时,主节点自动更新节点列表。当接收到用户访问副本的请求时,主节点会从存放副本的从节点中选择一个当前综合性能最好的节点作为最佳副本。当接收到用户创建副本的请求时,主节点会从从节点列表中选择一个当前性能最好的节点作为副本的存放位置。

3.2 数学描述

在系统评价中,我们通过综合考虑节点当前的CPU状态、内存状态、磁盘状态、网络状态等指标,来衡量节点的优劣。

云存储系统的数学描述如下:

数据节点集:设云存储系统中的数据节点集合 $D = \bigcup_{j=1,2,\dots,m} \{D_j\}$,其中 D_j 表示编号为 j 的数据节点, m 为数据节点的总数。

文件集:设文件集合 $F = \bigcup_{i=1,2,\dots,n} \{f_i\}$,其中 f_i 表示编号为 i 的文件, n 为文件的总数。

文件属性:将每个文件定义为四元组 $f_i = \langle n_i, u_i, s_i, h_i \rangle$,分别表示文件名、用户名、文件大小、文件热度。对文件 f_i 访问请求到达的频率服从Poisson分布。

数据节点属性:将云存储系统中的单个数据节点定义为七元组 $N_j = \langle CPU_j, Disk_j, Memory_j, IO_j, BWU_pj, BWDown_j, DiskFree_j \rangle$,分别表示CPU利用率、内存利用率、IO利用率、上传带宽利用率、下载带宽利用率、磁盘空闲空间。

请求集:设请求集 $R = \bigcup_{k=1,2,\dots,|R|} \{r_k\}$,其中 r_k 代表编号为 k 的请求, $|R|$ 表示请求的个数。

请求属性:将读写文件请求 r_k 定义为 $r_k = \bigcup_{k=1,2,\dots,|R|} \{fid_k\}$,其中 fid_k 是 r_k 所存储或访问的文件序列号。

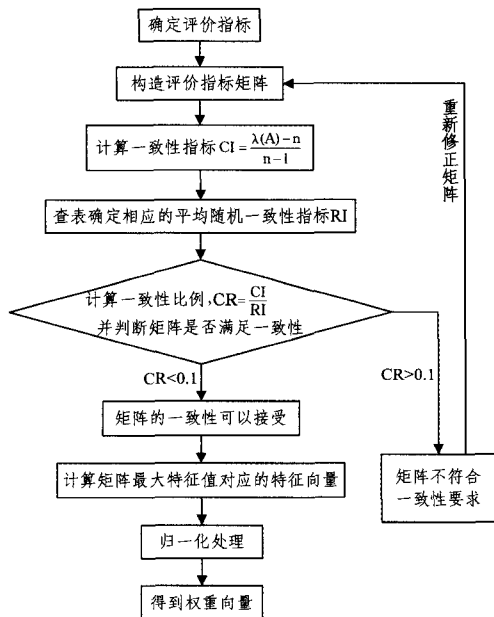


图2 AHP算法执行流程

在对上述多种指标进行加权评价处理时,本文选择层次分析法来实现各种指标从定性到定量的转化,得到指标所占的权重。

层次分析法^[16](Analytic Hierarchy Process, AHP)主要用于在多个影响因素的情况下评价各方案的优劣。它属于多属性决策中的重要方法,在关于离散的、有限个决策方法的决策问题上,这种方法比较适用。因此,本文通过定性描述各个指标的重要性,利用层次分析法将各个指标进行分层,利用数学方法得到各项指标的权值,完成从定性到定量的转化。AHP算法的执行流程如图2所示。

4 基于负载均衡的云存储环境下副本选择策略

云存储环境下,用户任务具有工作量大、随机性强等特点。副本管理策略中负载均衡是一个典型的NP困难问题,利用传统方法求得此类优化问题的解需要极大的时间消耗。因此,本文利用粒子群算法搜索文件副本最佳位置和访问方案,改进粒子群算法的计算步骤如下。

(1)初始化参数。设群体的规模为 N , N 为一批文件数,为每个文件随机指定一个合理的节点,则对应的矩阵为

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} & v_{11} & v_{12} & \dots & v_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} & v_{21} & v_{22} & \dots & v_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nn} & v_{N1} & v_{N2} & \dots & v_{Nn} \end{bmatrix}$$

其中, $\{x_{ij}, i=1,2,\dots,N, j=1,2,\dots,n\}$ 表示粒子 i 的位置为 j , v_{ij} 表示粒子 i 的速度, j 和 v_{ij} 为 $[0, m]$ 上满足均匀分布的随机数。

(2)计算粒子适应度。节点性能的加权评价适应度函数为 $F = D \times W$ 。其中, D 代表各个数据节点的各项指标:CPU、磁盘、I/O、内存、带宽; W 代表根据层次分析法得到的各项指标的权重。综合性能越高的节点,适应度越高。

(3)计算当前迭代中所有粒子的最优位置,即局部最优解。与全局最优解比较,更新全局最优解。

(4)更新粒子的速度和位置。改进PSO算法收敛速度的快慢由自适应惯性权重 ω 的非线性递减来控制,主要用来平衡算法的局部和全局探索能力。引入 ω 后的粒子迭代公式如下所示:

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 (pBest[i]^k - x_i^k) + c_2 r_2 (gBest^k - x_i^k)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

$$\omega = \omega_{min} + (\omega_{max} - \omega_{min}) * \sqrt{\frac{Iter - Iter_c}{Iter}}$$

其中, $Iter$ 为算法允许的迭代次数的最大值; $Iter_c$ 为算法执行的当前迭代次数; ω_{max} 为惯性权重 ω 的最大值,范围一般为 $0.8 \sim 1.2$; ω_{min} 为惯性权重 ω 的最小值,通常取值为 0.4 ; c_1, c_2 为学习因子; r_1, r_2 为 $[0, 1]$ 内满足均匀分布的随机数。

(5)判断迭代是否终止,目标函数的适应度达到指定阈值或者当前迭代次数达到指定代数则终止,否则返回步骤(2)。

5 实验及结果分析

本文针对副本管理问题设计并实现了一个基于Java的云存储系统仿真平台,模拟从用户提交任务请求、副本选择策略执行到副本传输至目标节点的整个数据存储访问流程。为了进一步检验改进算法的真实效果,本文结合层次分析法和粒子群算法进行了另外3组对比仿真实验,并分析了在不同用户需求状况下3种方案各自的优势。

5.1 实验环境与配置

本文实验运行在自主搭建的小型云平台中,由一台主节点和若干台可动态扩展的存储节点构成。仿真平台由以下5大模块组成。

伪负载生成器模块:负责为副本选择算法生成访问数据的请求,模拟大量的访问请求。

资源及任务分析模块:负责接收用户的数据访问请求,以及接收并解析从节点发送过来的硬件性能信息和运行状态的动态信息。

副本选择策略模块:根据接收到的用户请求以及收集到的云存储平台的资源信息,结合本文提出的副本选择算法模拟执行实际数据中心的副本选择策略。

传输数据模块:根据副本选择策略模块得到的结果,完成用户访问请求。

结果显示模块:输出副本选择策略得到的最佳副本位置以及算法消耗时间、各项指标的加权值等相关实验数据。

为了更详细具体地分析本文的副本选择策略的实际效果,基于用户偏好设计了另外3种副本选择策略,意义在于不同用户的需求不同,因此对系统的性能要求也有偏差。给出3种副本选择策略的详细描述如下:

(1)最快传输速度策略(MFTS)

传输速度主要与带宽相关,另外磁盘性能也在一定程度上影响传输速度。因此,在这种策略中,侧重考虑当前节点的带宽利用率和磁盘I/O利用率,在层次分析法中提高带宽和I/O的重要性。见评价矩阵 A_{MFTS} 。对应的指标为CPU、内存、磁盘空间、I/O、带宽。

(2)最高安全性策略(MHS)

安全性是一些云用户的特殊需要,如企业中的重要数据需要较高的安全性和私密性,因此,在这种策略中,侧重考虑当前节点的安全性,在层次分析法中提高安全性相比其他指标的重要性。见评价矩阵 A_{MHS} 。对应的指标为安全等级、内存、磁盘空间、I/O、带宽。

(3)最小存储空间成本策略(MSSC)

基于最小成本的副本管理策略是当前该领域中的一个重点研究方向。这种策略可以很大程度上为企业和云提供商降低冗余备份策略的成本。在这种策略中,侧重考虑磁盘空间利用率,在层次分析法中提高磁盘空间利用率相比于其他指标的重要性。见评价矩阵 A_{MSSC} 。对应的指标为CPU、内存、磁盘空间、I/O、带宽。

通过分析节点性能的影响度,结合本文实验应用场景,为使系统一定程度上达到负载均衡,得到如下矩阵 A ,并且设计另外3个对比组。

本文实验中使用的其中一组评价矩阵如下所示:

$$A = \begin{bmatrix} 1 & 6 & 6 & 9 & 9 \\ \frac{1}{6} & 1 & \frac{1}{3} & 5 & 5 \\ \frac{1}{6} & 1 & 1 & 6 & 6 \\ \frac{1}{9} & \frac{1}{5} & \frac{1}{6} & 1 & 1 \\ \frac{1}{9} & \frac{1}{5} & \frac{1}{6} & 1 & 1 \end{bmatrix}$$

$$A_{MFTS} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{8} & \frac{1}{9} \\ 2 & 1 & \frac{1}{3} & \frac{1}{6} & \frac{1}{7} \\ 5 & 3 & 1 & \frac{1}{4} & \frac{1}{5} \\ 8 & 6 & 4 & 1 & \frac{1}{2} \\ 9 & 7 & 5 & 2 & 1 \end{bmatrix}$$

$$A_{MSSC} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{9} & \frac{1}{8} & \frac{1}{6} \\ 2 & 1 & \frac{1}{7} & \frac{1}{6} & \frac{1}{4} \\ 9 & 7 & 1 & 3 & 5 \\ 8 & 6 & \frac{1}{3} & 1 & 4 \\ 6 & 4 & \frac{1}{5} & 4 & 1 \end{bmatrix}$$

$$A_{MHS} = \begin{bmatrix} 1 & 9 & 6 & 8 & 4 \\ \frac{1}{9} & 1 & \frac{1}{5} & \frac{1}{2} & \frac{1}{8} \\ \frac{1}{6} & 5 & 1 & 2 & \frac{1}{5} \\ \frac{1}{8} & 2 & \frac{1}{2} & 1 & \frac{1}{7} \\ \frac{1}{4} & 8 & 5 & 7 & 1 \end{bmatrix}$$

5.2 实验结果及分析

本节通过模拟不同用户需求不同任务数量下副本策略的任务执行情况,给出4种副本选择策略的执行结果,对各副本策略下存储空间利用率、CPU占用率、内存占用率和I/O占用率进行比较。经过多次实验得到如图3所示的平均消耗时间的对比。

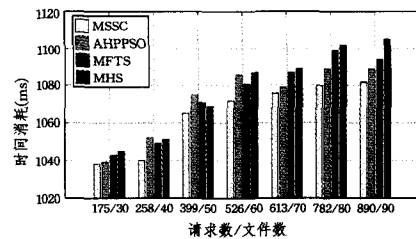


图3 4种策略在时间消耗上的对比

在图4中,随着用户访问量的增加,任务数量增加,CPU利用率先呈逐渐上升的趋势,然后一段时间内在10%上下浮动,最后随着任务逐渐处理完成,任务数量减少,CPU利用率呈明显下降趋势。

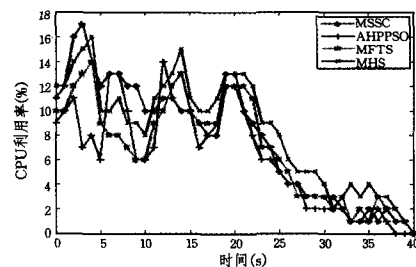


图4 CPU占用率

在图5中,随着用户的并发访问量增加,存取副本的频次增加,直接造成资源利用率的攀升。随着任务量下降,存储资

源利用率呈逐渐下降趋势。其中, MSSC 策略在整个当前一批任务执行阶段都能获得最高的资源利用率。特别是在任务数量减少时, MSSC 策略在存储资源利用率方面的优势更加明显。

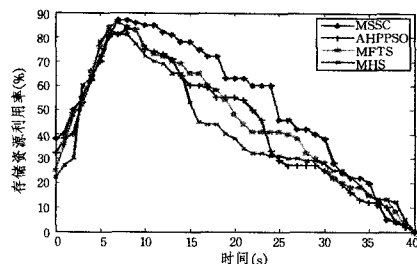


图5 存储资源利用率

在图6中,用户的并发访问量逐渐增加导致网络上传送的数据量增加,直接造成带宽利用率先呈上升趋势,当并发访问量上下波动时,带宽利用率也因此呈现先上升后下降的趋势。其中, MFTS 策略在整个过程中都能获得较低的带宽利用率,即较优等的网络。由此可见, MFTS 在传输速度方面具有较好的性能。

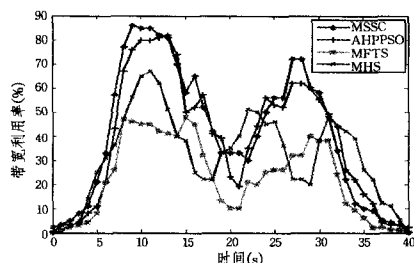


图6 带宽利用率

在图7中,随着任务数量逐渐增加,内存利用率首先呈逐渐上升的趋势,然后一段时间内趋于稳定,最后随着任务数量下降,内存利用率呈下降趋势,直到所有任务都处理完成。

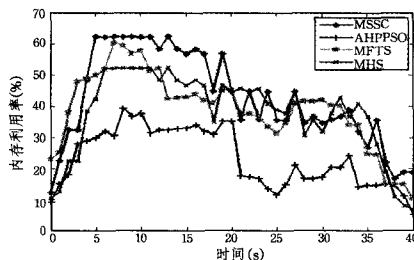


图7 内存占用率

本文提出的基于负载均衡的副本管理策略 AHPPSO 在处理任务过程中可以获得较低的 CPU 占用率、内存占用率、存储资源利用率和相对中等水平的带宽利用率。可见,这种策略一定程度上平衡了系统的负载状况,得到了较均衡的系统总体性能。

综上所述,对于不同的用户需求,各个副本策略可以根据相应的指标动态地选择所需副本,更好地利用云平台的资源。在保证数据可靠性的同时,各个方案在其自身优化目标方面有着明显优势。

结束语 云提供商的用户群体众多,需求多样,如何为用户提供满意的、合适的服务是云提供商关心的重点。在不增加存储服务成本的同时,依据各类用户需求的差异,本文提出

了一种基于用户需求差异化的云提供商数据可靠性保证算法 AHPPSO。AHPPSO 以满足用户个性化需求为目标,在一定用户任务访问量下进行对比实验。分别针对缩短用户响应时间、保证数据高安全性和节约存储成本方面设计了3种副本选择方案。实验结果证明:3种副本选择策略在保证系统负载均衡的情况下,在各自优化目标方面性能得到很大的提高。

参考文献

- [1] Josef S, Johannes M, Alexander S. Creating optimal cloud storage systems [J]. *Future Generation Computer Systems*, 2013, 29(4):1062-1072
- [2] 李建江,崔健,王聃,等. MapReduce 并行编程模型研究综述[J]. *电子学报*, 2011, 39(11):2636-2643
- [3] Zhang Da-wei, Sun Fu-quan, Cheng xu, et al. Research on Hadoop-based enterprise file cloud storage system[C]// *Proceedings of 2011 3rd International Conference on Awareness Science and Technology (iCAST2011)*. Dalian, China: IEEE Computer Society, 2011:434-437
- [4] Wei Qing-song, Bharadwaj V, Gong Bo-zhao, et al. CDRM: a cost-effective dynamic replication management scheme for cloud storage cluster [C]// *Proceedings of 2010 IEEE International Conference on Cluster Computing*. Heraklion, Crete, Greece: Institute of Electrical and Electronics Engineers Inc, 2010:188-196
- [5] Li Wen-hao, Yang Yun, Chen Jin-jun. A cost-effective mechanism for Cloud data reliability management based on proactive replica checking[C]// *Proceedings of 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*. 2012
- [6] Lin Jenn-wei, Chen Chien-hung, Chang J M. QoS-Aware Data Replication for Data-Intensive Applications in Cloud Computing Systems[C]// *Proceedings of IEEE Transactions on Cloud Computing*. 2013
- [7] 程振东,栾钟治,孟由,等. 云文件系统中纠删码技术的研究与实现[J]. *计算机科学与探索*, 2013(4):315-325
- [8] 付园. 基于 HDFS 的优化数据冗余策略的研究[D]. 吉林:吉林大学, 2014
- [9] 罗象宏,舒继武. 存储系统中的纠删码研究综述[J]. *计算机研究与发展*, 2012, 49(1):1-11
- [10] 黑继伟. 基于分布式并行文件系统 HDFS 的副本管理模型[D]. 吉林:吉林大学, 2010
- [11] 娄超. 云存储环境中副本选择与一致性维护策略研究[D]. 济南:山东师范大学, 2014
- [12] 杜芸芸. 一种面向纠删码技术的云存储可靠性机制[J]. *计算机应用与软件*, 2014, 31(2):312-316, 320
- [13] 杨东日,王颖,刘鹏. 一种副本复制和纠错码融合的云存储文件系统容错机制[J]. *清华大学学报(自然科学版)*, 2014, 54(1):137-144
- [14] Xu zhi-hong, Hou Xiang-dan, Sun Ji-zhou. Ant Algorithm-based Task Scheduling[C]// *Grid Computing Electrical and Computer Engineering*. 2003:1107-1110
- [15] 张新亮,万晓冬. 基于遗传算法的副本管理策略[J]. *计算机仿真*, 2009, 26(1):197-199, 263
- [16] 张兴. 基于 Hadoop 的云存储平台的研究与实现[D]. 成都:电子科技大学, 2013