

先进计算

大规模云存储系统副本布局研究

董继光,陈卫卫,田浪军,吴海佳

解放军理工大学 指挥自动化学院,南京 210007

摘要: 在基于副本冗余机制的大规模云存储系统中,以往的副本布局算法只能部分地满足副本布局中高可靠、高可扩展并且高效的要求,针对这一问题,提出了一种基于分组和一致性哈希的副本布局算法。首先,将关联性高的存储节点进行分组;然后,通过改进的一致性哈希算法将同一数据对象的多个副本分别分配到不同分组中;最后,再通过改进的一致性哈希算法将分配到各分组的数据副本放置在组内对应的存储节点上。理论分析可知,该方法大大提高数据的可靠性。仿真结果表明,该算法能满足副本布局的均衡性、自适应性要求,并能在几十微秒内完成副本定位。

关键词: 云存储 副本布局 一致性哈希算法 分组 可靠性

Replica placement study in large-scale cloud storage system

DONG Ji-guang, CHEN Wei-wei, TIAN Lang-jun, WU Hai-jia

Institute of Command Automation, PLA University of Science and Technology, Nanjing Jiangsu 210007, China

Abstract: In the large-scale cloud storage system based on copy redundancy, previous layout algorithm can only partially meet the requirements of high reliability, high scalability and high efficiency in the replica layout. To solve this problem, this paper proposed a Replica Placement algorithm based on Grouping and Consistent Hashing (RPGCH). The storage nodes were classified into different groups by their correlativity, then the replicas of one object were assigned in different groups by consistent hashing algorithm, after that each replica was placed into corresponding storage node in the group by consistent Hashing algorithm. The theoretical analysis proves that the reliability of data is improved. The simulation results show that RPGCH can assign data evenly among storage nodes and adapt well with the changing scale of cloud storage system. Moreover, RPGCH is time efficient with little memory overhead.

Keywords: cloud storage replica layout consistent hashing algorithm grouping reliability

收稿日期 2011-09-19 修回日期 2011-11-16 网络版发布日期 2012-03-01

DOI:

基金项目:

国家自然科学基金资助项目(60603029);国家863计划项目(2008AA01A309)。

通讯作者: 董继光

作者简介: 董继光(1986-),男,河南周口人,硕士研究生,CCF会员,主要研究方向:分布式网络存储、云存储;陈卫卫(1967-),女,四川隆昌人,教授,CCF会员,主要研究方向:软件工程、云计算、分布式网络存储;田浪军(1985-),男,陕西渭南人,硕士研究生,CCF会员,主要研究方向:分布式数据库;吴海佳(1986-),男,江苏南通人,博士研究生,CCF会员,主要研究方向:分布式网络存储、分布式文件系统。

作者Email: pladjg@163.com

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(814KB)
- ▶ [HTML全文]
- ▶ 参考文献[PDF]
- ▶ 参考文献

服务与反馈

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ 引用本文
- ▶ Email Alert
- ▶ 文章反馈
- ▶ 浏览反馈信息

本文关键词相关文章

- ▶ 云存储
- ▶ 副本布局
- ▶ 一致性哈希算法
- ▶ 分组
- ▶ 可靠性

本文作者相关文章

- ▶ 董继光
- ▶ 陈卫卫
- ▶ 田浪军
- ▶ 吴海佳

PubMed

- ▶ Article by Dong,J.G
- ▶ Article by Chen,W.W
- ▶ Article by Tian,L.J
- ▶ Article by Wu,H.J

参考文献:

[1]GRAY J. What next? A few remaining problems in information technology[EB/OL]. [2011-05-10].

[2]云存储[EB/OL]. [2011-05-10]. <http://baike.baidu.com/view/2044736.htm>.

[3]GHEMAWAT S, GOBIOFF H. The Google file system [C]// Proceedings of the 19th ACM Symposium on Operating Systems Principles. New York: ACM Press, 2003: 19-22.

[4]SHVACHKO K, KUANG H, RADIA S. The Hadoop distributed file system [C]// IEEE 26th Symposium on Storage Systems and Technology. Piscataway, NJ: IEEE Press, 2010: 1-10.

[5]WEIL S A, BRANDT S A, MILLER E L, et al. Ceph: A scalable, high-performance distributed file system [EB/OL]. [2010-05-10]. <http://www.ssrc.ucsc.edu/Papers/weil-osdi06.pdf>.

[6]KARGER D, LEHMAN E, LEIGHTON T, et al. Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web [C]// STOC'97: Proceedings of the 29th Annual ACM Symposium on Theory of Computing. New York: ACM Press, 1997: 654-663.

[7]BRINKMANN A, EFFERT S, auf der HEIDE F M. Dynamic and redundant data placement [C]// ICDCS'07: Proceedings of 27th International Conference on Distributed Computing Systems. Piscataway, NJ: IEEE Press, 2007: 29.

[8]HONICKY R J, MILLER E L. Replication under scalable hashing: a family of algorithms for scalable decentralized data distribution [C]// Proceedings of the 18th International Parallel and Distributed Processing Symposium. Piscataway, NJ: IEEE Press, 2004: 96.

[9]WEIL S A, BRANDT S A, MILLER E L, et al. CRUSH: Controlled, scalable and decentralized placement of replicated data [C]// Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. New York: ACM Press, 2006: 31-42.

[10]XIAO N, CHEN T. RAEDP: An effective hybrid data placement algorithm for large-scale storage systems [J]. Journal of Supercomputing, 2011,55(1): 103-122.

[11]DeCANDIA G, HASTORUN D. Dynamo: Amazon's highly available key-value store [C]// Proceedings of the 21st ACM SIGOPS Symposium on Operating Systems Principles. New York: ACM Press, 2007: 14-17.

#### 本刊中的类似文章

1. 周敬利 周正达.改进的云存储系统数据分布策略[J]. 计算机应用, 2012,32(02): 309-312
2. 曹夕 许力 陈兰香.云存储系统中数据完整性验证协议[J]. 计算机应用, 2012,32(01): 8-12
3. 苏崇茂.7轮ARIA-256的不可能差分新攻击[J]. 计算机应用, 2012,32(01): 45-48
4. 蔡植善.基于通用分组无线业务短信控制的LED点阵屏设计[J]. 计算机应用, 2011,31(12): 3449-3452
5. 郑洪英 李文杰 肖迪.基于时空混沌系统的图像分组加密算法[J]. 计算机应用, 2011,31(11): 3053-3055
6. 葛君伟 李志强 方义秋.云存储环境下基于分散式服务器的Erasure Code算法[J]. 计算机应用, 2011,31(11): 2940-2942
7. 吕堂祺 黄宁 贾晓光 王东.面向服务架构软件实现前的可靠性评价方法[J]. 计算机应用, 2011,31(09): 2436-2439
8. 刘立群.集中式无线局域网分离介质访问控制的CCMP设计[J]. 计算机应用, 2011,31(08): 2159-2161
9. 胡亚明 邓亚平 杨佳.负载均衡的无线传感器网络自适应分组成簇算法[J]. 计算机应用, 2011,31(08): 2056-2058
10. 张晓南 刘安心 刘斌 张宏梅 青星.基于优化PSO-SVM模型的软件可靠性预测方法[J]. 计算机应用, 2011,31(07): 1762-1764
11. 李强 李城鑫 黄玉清 姚远程.相关信道下基于多用户STBC-OFDM系统的子载波分配研究[J]. 计算机应用, 2011,31(07): 1948-1951
12. 韩炫 雷航.考虑环境因素的软件可靠性增长模型[J]. 计算机应用, 2011,31(07): 1759-1761
13. 邓亚平 陈峥.能量负载均衡的无线传感网分组成簇协议[J]. 计算机应用, 2011,31(06): 1465-1468
14. 沈雅芬 黄宁 彭永义.OWL-S模型转化为重写逻辑模型的方法[J]. 计算机应用, 2011,31(06): 1491-1494
15. 卢丹华 钟诚 杨锋.基于多核多线程的AES保密模式[J]. 计算机应用, 2011,31(04): 1003-1005
16. 孟江涛 卢显良.虚拟机监控器Xen的可靠性优化[J]. 计算机应用, 2010,30(9): 2358-2361
17. 孟海宁 刘建军.基于粗糙小波网络应用服务器老化预测模型[J]. 计算机应用, 2010,30(8): 2024-2028

18. 李权 鄢萍 贺晓辉.高可靠性的嵌入式软件现场更新方法[J]. 计算机应用, 2010,30(8): 2231-2231
19. 周鑫 张锦 赵研科 王如龙.GPRS小区流量预测中时序模型比较研究[J]. 计算机应用, 2010,30(4): 884-887
20. 楼俊钢 江建慧.软件失效时间数据相关性研究[J]. 计算机应用, 2010,30(3): 600-602
21. 杨宏志 韩文报 赵龙.适于硬件实现的S盒构造方法[J]. 计算机应用, 2010,30(3): 674-676
22. 蔡少杰 林亚平 易叶青 叶松涛.无线传感器网络基于分组协商的数字水印算法[J]. 计算机应用, 2010,30(3): 688-691
23. 于建志 陈永生.磁浮列车自动驾驶系统控制策略及可靠性研究[J]. 计算机应用, 2010,30(12): 3419-3422
24. 相丽 潘峰 钮可 郭耀.基于运动矢量的视频隐写方法[J]. 计算机应用, 2010,30(11): 3022-3024
25. 林志雄 鄢萍 贺晓辉.面向嵌入式无线信息终端的在线升级方法[J]. 计算机应用, 2010,30(07): 1725-1727
26. 刘唐 彭舰.P2P文件共享系统中的分组信誉驱动机制[J]. 计算机应用, 2010,30(07): 1789-1793
27. 晁冰 徐仁佐.基于LSSVRM和SA算法的软件失效模型[J]. 计算机应用, 2010,30(06): 1648-1650
28. 王雪飞.MP\_WSMN数据融合权重 $\mu$ - $\xi$ 的高QoS实现[J]. 计算机应用, 2010,30(06): 1469-1471
29. 刘丹 谢文君.P2P网络下的KNN查询[J]. 计算机应用, 2010,30(05): 1156-1158
30. 王晓原 吴芳 邢丽.交通流冗余数据识别和约简方法[J]. 计算机应用, 2009,29(4): 1110-1113
31. 靳昂 江建慧 楼俊刚 张蕊.基于GM模型的软件可靠性建模[J]. 计算机应用, 2009,29(3): 690-694
32. 张立鹏 顾华玺 王长山.一个新颖的模块化NoC路由器设计[J]. 计算机应用, 2009,29(2): 511-513
33. 张红霞 戴居丰.3GPP信道模型中STBC-OFDM系统性能分析[J]. 计算机应用, 2009,29(10): 2614-2616
34. 张宏斌 贾志新.基于逆向思维的系统可靠性精确分配模型[J]. 计算机应用, 2009,29(09): 2468-2470
35. 满梦华 原亮 丁国良 巨政权 宋亮.嵌入式高可靠性异构双机冗余系统的设计[J]. 计算机应用, 2009,29(08): 2143-2145
36. 姚雪芬 黎福海.结合子载波分群检测方法的组天线MIMO-OFDM系统研究[J]. 计算机应用, 2009,29(08): 2157-2160
37. 赵文栋 彭来献 田畅.传感器反应网络协作机制综述[J]. 计算机应用, 2009,29(08): 2183-2187
38. 杨宏志 韩文报.一类分组密码的S盒重组算法[J]. 计算机应用, 2009,29(08): 2198-2199
39. 刘永 李言 李淑娟 侯晓莉 曾志斌.基于可靠性分析的MAS构建优化算法[J]. 计算机应用, 2009,29(06): 1532-1535
40. 吕宁 孙广明 张宇.基于多混沌系统的图像分组密码设计[J]. 计算机应用, 2008,28(9): 2263-2266
41. 周娜琴 张友生.构件软件的可靠性估算模型[J]. 计算机应用, 2008,28(6): 1630-1631
42. 陈逸菲 张颖超 叶小岭.关系数据库中分组查询的模糊扩展及去模糊机制[J]. 计算机应用, 2008,28(5): 1177-1179
43. 吴宇 魏急波 习勇.一种针对混合的实时/非实时业务的无线调度算法[J]. 计算机应用, 2008,28(2): 389-393
44. 高洁 袁家斌 徐涛 齐艳珂.一种基于混合反馈的混沌图像加密算法[J]. 计算机应用, 2008,28(2): 434-436
45. 滕云龙 师奕兵 康荣雷.软件可靠性组合预测模型研究[J]. 计算机应用, 2008,28(12): 3092-3094
46. 田树华 宋立明.一种基于竞争层神经网络的新型图像轮廓线分组算法[J]. 计算机应用, 2008,28(12): 3147-3149
47. 何小卫 王爱华 马跃.基于GPRS的GPS车载终端通信技术研究[J]. 计算机应用, 2008,28(11): 2952-2954
48. 刘仲 李宗伯 阳柳.基于对象分组的集群存储系统可靠性分析模型[J]. 计算机应用, 2008,28(1): 36-38
49. 许兰 朱巧明 李培峰 贡正仙.一种基于历史信息反馈的动态服务资源选择模型[J]. 计算机应用, 2007,27(9): 2284-2286
50. 费岚 潘春建 谭红艳.基于业务流均衡的802.16无线分组调度算法[J]. 计算机应用, 2007,27(8): 1865-1867
51. 汤念 王雷 姚焯善 张大方 徐红云.一种基于分组填充Mix策略的匿名通信机制[J]. 计算机应用, 2007,27(7): 1606-1608
52. 郭健彬 曾声奎 陈云霞.基于STEP的RMS信息模型研究及实现[J]. 计算机应用, 2007,27(6): 1445-1447
53. 余愚 邓志平.动态数组在系统可靠性与维修性仿真中的应用[J]. 计算机应用, 2007,27(6): 1521-1523
54. 李宝林 李志蜀 陈良银 倪建成.基于测试用例的软件可靠性分析[J]. 计算机应用, 2007,27(5): 1254-1256
55. 谢文君 徐静 吴严寒.一种基于Chord优化的空间数据存储方法[J]. 计算机应用, 2007,27(3): 528-530
56. 许宁 张毅坤.基于正交分层聚类算法软件可靠性模型的预测分析[J]. 计算机应用, 2007,27(3): 635-637
57. 李薇 张凤鸣.发布订阅系统在移动Ad Hoc网络中的三种路由算法[J]. 计算机应用, 2007,27(3): 514-515

58. 原菊梅; 侯朝楨; 王小艺; 高琳.基于包含度理论的软件可靠性多模型综合动态预计[J]. 计算机应用, 2007,27(2): 387-388
59. 何跃; 杨剑; 徐玖平.基于GMDH的组合预测模型应用研究[J]. 计算机应用, 2007,27(2): 456-458
60. 钟黔川; 朱清新.Blowfish密码系统分析[J]. 计算机应用, 2007,(12): 2940-2941
61. 叶益群; 梁华国; 祝沈财; 张念; 詹凯华.分组频率Golomb码测试数据压缩研究[J]. 计算机应用, 2007,(12): 3136-3138
62. 周翔翔; 尹忠海; 刘守义; 韩毅娜.一种基于密钥的水印嵌入位置置乱算法[J]. 计算机应用, 2007,27(10): 2473-2474
63. 吴玉美; 阮镰.软件可靠性测试的加速机理研究[J]. 计算机应用, 2006,26(6): 1449-1451
64. 任泰云; 候红.基于CORBA的通用无线分组业务远程监控系统[J]. 计算机应用, 2006,26(6): 1308-1310
65. 刘义; 王玲; 刘辉.正交空时分组码系统的一种新的盲信道估计算法[J]. 计算机应用, 2006,26(12): 2793-2795
66. 赵伟; 莫国庆; 那宝玉; 刘鹏.基于Reed-Solomon算法的RAID机制的设计与实现[J]. 计算机应用, 2006,26(11): 2756-2758
67. 王亚平; 周兴社; 闫守孟; 张凡.ESSDRR调度算法的性能仿真与分析[J]. 计算机应用, 2005,25(12): 2904-2907
68. 吴勤; 侯朝楨; 原菊梅.基于Kohonen网络的软件可靠性模型选择[J]. 计算机应用, 2005,25(10): 2331-2333
69. 石贤良; 吴成富.双冗余飞行控制计算机系统设计及实现[J]. 计算机应用, 2005,25(09): 2163-2164
70. 吴娟; 马永强; 刘影.一种基于主备机快速切换的双机容错系统[J]. 计算机应用, 2005,25(08): 1948-1951
71. 张树旗; 贾树恒.一种支持变长分组的CIOQ交换结构[J]. 计算机应用, 2005,25(07): 1491-1493
72. 尚玲; 陈峰; 徐文立.基于动态分组的三维小波变换方法[J]. 计算机应用, 2005,25(05): 1119-1121
73. 刘航; 戴冠中; 李晖晖; 慕德俊.工作于CBC模式的AES算法可重配置硬件实现[J]. 计算机应用, 2005,25(01): 135-137