中国科学院研究生院学报 2011, 28(6) 793-800 DOI: ISSN: CN:

论文

一种基于模糊推理的Hadoop异构机群自动配置工具

代栋, 周学海, 杨峰, 王超

中国科学技术大学苏州研究院嵌入式系统实验室, 江苏苏州 215123; 中国科学技术大学计算机科学与技术学院, 合肥 230027

摘要:

以运行Hadoop云计算软件平台的异构机群中服务器的硬件指标作为模糊输入,设计了一种自动配置本地服务器的工具.该工具改变了传统云计算平台下的配置方法,以服务器运行历史数据作为参考,引入异构配置方法,通过对异构机群特性的分析,实现对机群的自动配置.仿真实验证明,在保证机群高效的基础上,该工具降低了云计算异构机群的维护成本,具有较高的通用性和可扩展性,对一般的云计算平台也具有参考价值.

关键词: Hadoop 自动配置 模糊逻辑 模糊推理

An auto-configuration tool for heterogeneous Hadoop cluster

DAI Dong, ZHOU Xue-Hai, YANG Feng, WANG Chao

Suzhou Advanced Institution of USTC, Embedded System Lab, Suzhou 215123, Jiangsu, China; Computer Science and Technology College, USTC, Hefei 230027, China

Abstract:

The rapid development of cloud computing makes the heterogeneous cluster a hot research topic. One of the urgent problems in this field is how to configure the cloud computing software platform running on a heterogeneous cluster. We propose a new tool, which can configure each local server automatically according to its hardware parameters in a Hadoop cluster and configures each server individually by analyzing the heterogeneous features and the running history of the Hadoop cluster instead of the traditional uniform way. The simulation and experiment results show that this tool can obviously reduce the maintenance cost without any degradation of system performance and improve performance compared to the manual optimization. It is believed that this method is general and extensible and should be valuable for other cloud computing platforms.

Keywords: Hadoop auto-configure fuzzy logic fuzzy inference

收稿日期 2010-07-19 **修回日期** 2010-09-06 **网络版发布日期**

DOI:

基金项目:

国家自然科学基金(60873221)及思科大学研究计划资助

通讯作者:

作者简介:

作者Email: daidong@mail.ustc.edu.cn

参考文献:

[1] Apache Software Foundation. The apache hadoop project.http://hadoop.apache.org/, as of 15/06/2009.

[2] Wikipedia. Heterogeneous_network.http://en.wikipedia.org/wiki/Heterogeneous_network.

[3] Bhandarkar M, Gogate S, Bhat V. Hadoop performance tuning: a case study. http://cloud.citris-uc.org/system/files/private/BerkeleyPerformanceTuning.pdf.

[4] Hadoop cluster setup.http://hadoop.apache.org/common/docs/current/cluster_setup.html.

[5] 闻新, 周露, 李东江, 等. MatLab模糊逻辑工具箱的分析与应用
[M]. 北京:科学出版社, 2001.

[6] Dean J, Ghemawat S. Mapreduce: simplified data processing on large clusters //Proc of OSDI. 2004: 137-150.

[7] Chang F, Dean J, Ghemawat S, et al. Bigtable: a distributed storage system for structured data [J]. ACM Trans Comput Syst, 2008,26(2): 1-26.

[8] Boulon J, Konwinski A, Qi R, et al. Chukwa, a large-scale monitoring system //Cloud Computing and its Applications. Chicago, IL, October 2008:1-5.

[9] Thusoo A, Sarma J S, Jain N, et al. Hive-A warehousing solution over a map-reduce framework [J]. PVLDB, 2009, 2(2):1626-1629.

[10] Hadoop. Powered by Hadoop. http://wiki.apache.org/hadoop/PoweredBy.

[11] Murthy A C. Speeding up Hadoop.
http://developer.yahoo.com/blogs/ydn/posts/2009/09/hadoop_summit_speeding_up_hadoop/.

[12] Sharma S. Advanced Hadoop tuning optimization. http://www.slideshare.net/ImpetusInfo/ppt-on-advanced-hadoop-tuning-n-optimisation.

[13] Ghemawat S, Gobioff H, Leung S. The Google file system //Proceedings of the 19th ACM Symposium on Operating Systems Principles 2003. SOSP, 2003:29-43. DOI: 10.1145/945445.945450, URL:http://portal.acm.org/citation.cfm?id=945445.945450.

[14] Kapadia N H, Fortes J A B, Brodley C E. Predictive application-performance modeling in a computational grid environment //Proceedings of the IEEE International Symposium on High Performance Distributed Computing 1999. 1999. DOI: 10.1109/HPDC.1999.805281.

[15] Wang G, Butt A R, Pandey P, et al. A simulation approach to evaluating design decisions in MapReduce setups //Proceedings of the 17th Annual Meeting of the IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems, MASCOTS 2009. DOI: 10.1109/MASCOT.2009.5366973.2009: 1-11.

[16] Malley O O, Murthy A C. Winning a 60 second dash with a yellow elephant . TR, Yahoo! Inc, 2009.

本刊中的类似文章