

基于 Poisson 过程的 Hadoop 集群平均寿命计算方法

李佳宣, 郑晓薇, 张 永

(辽宁师范大学 计算机与信息技术学院, 辽宁 大连 116081)

摘 要: Hadoop 集群在连续长时间执行任务的过程中, 集群故障会对集群造成不同程度的冲击与损害, 是集群性能下降、系统故障频发的主要原因, 更会导致集群平均寿命的大大缩短. 为解决此问题, 提出了一种基于 Poisson 过程的 Hadoop 集群平均寿命预测方法. 通过预测计算, 及时对集群进行调整优化, 使系统性能保持在正常状态. 同时在三种不同特征类型的 Hadoop 集群的实验中观察集群在任务执行过程中平均寿命的变化. 实验结果表明, 该方法能有效地判断 Hadoop 集群故障发生的频率, 预测出集群的平均寿命, 对集群性能优劣进行评价并及时对集群性能进行优化

关键词: 集群故障; Poisson 过程; 集群平均寿命

中图分类号: TP302

文献标识码: A

文章编号: 1000-7180(2016)07-0150-04

A Calculation Method of Hadoop Cluster Average Life Based on Poisson Process

LI Jia-xuan, ZHENG Xiao-wei, ZHANG Yong

(College of Computer and Information Technology, Liaoning Normal University, Dalian 116081 China)

Abstract: After a long task execution on Hadoop cluster, as cluster faults would hurt Hadoop cluster on different strength, which became the main reason of cluster performance degradation and more frequent cluster faults, more serious would cause the decrease of cluster average life. To solve the problem, this paper proposes a method of Hadoop cluster average life prediction based on Poisson process. By prediction calculation, adjust and optimize the cluster timely to keep the cluster performance in a normal state. Then make several experiments on three different feature types of Hadoop cluster to observe the frequency of cluster average life. According to the experiment results, this method is the significant to Hadoop cluster performance evaluation and optimization.

Key words: cluster faults; poisson process; cluster average life

1 引言

在云计算技术迅猛发展的今天, 拥有 MapReduce 计算架构的 Hadoop 集群作为云计算技术的主要操作平台, 在互联网搜索、数据挖掘、移动通信等许多领域中得到了广泛应用^[1-3]. 近期随着大数据技术的提高, 多种新型计算架构也不断涌现, 如 Yarn、Spark 等. 但新的技术在起步的初期阶段仍然面临着许多问题亟待解决, 相应的应用开发还很薄弱. 绝大多数的产品与应用还是将成熟的 Hadoop 集群作

为其最重要计算平台, 因此对于 Hadoop 集群问题的研究仍然具有重要的价值. Hadoop 集群可以由数目庞大的普通商用服务器节点构成, 该集群在提供大规模分布式数据存储资源的同时, 也提供了大规模的并行化计算资源, 具有成本低且性能高的巨大优势, 是许多商业应用的首选. 然而, 低成本的节点服务器在长时间连续执行任务的过程中易出现集群故障, 而每次故障的发生均会对 Hadoop 集群造成不同程度的冲击或损害, 导致集群平均寿命不断降低. 长此以往会出现 Hadoop 集群性能下降, 甚

收稿日期: 2015-09-18; **修回日期:** 2015-11-10

基金项目: 国家自然科学基金项目(61373127)

至集群崩溃。为了降低和避免这一现象发生的概率,并根据当前的集群性能状态做出及时的调整和优化,本文提出利用 Poisson 过程预测集群故障发生的概率,计算 Hadoop 集群的平均寿命,并将集群平均寿命作为评价集群性能优劣的标准,对改善集群性能有着十分重要的意义。

Poisson 过程是随机过程中一类重要的计数过程,在实际中有着广泛的应用,如电子媒体类^[4]、图形图像类^[5]等。由于 Poisson 过程具有平稳独立增量,过程在任何时刻都“重新开始”,具有“无记忆性”,其与指数分布的“无记忆性”是对应的。根据小概率事件原理,Hadoop 集群在一定时间内发生故障的概率是很小的,但考虑多个时间段的累计,故障的发生将会有一个大体稳定的速率,进而可以预测出集群未来性能情况。通过对符合 Poisson 分布的 Hadoop 集群故障统计过程的计算与推导可达到预测 Hadoop 集群平均寿命的目的,经实验验证获得较好的结果。

2 集群故障与集群损害

2.1 集群故障的 Poisson 分布

一般情况下,Hadoop 集群在执行任务过程中可分为两种状态:一种是集群正常运行状态,另外一种就是集群故障。

定义 1 集群故障。Hadoop 集群在执行任务过程中会出现导致集群效率下降或集群任务执行失败的事件,此类事件统称为集群故障。

较为普遍的 Hadoop 集群故障主要分为以下几种:节点失效、网络阻塞、任务异常和实际任务执行时间过长。集群故障的发生会对集群造成冲击或损害,使集群性能明显降低。

设集群故障记为事件 F ,其发生的次数为 λ ($\lambda > 0$), $F(t)$ 表示到时刻 t 被记录下来的事件总数,则 $F(t)$ 是一个具有平稳独立增量的计数过程,且 $F(0) = 0$ 。若满足 Poisson 过程的条件,则在时间一定的情况下集群故障的发生总次数 $F(t)$ 服从均值为 λt 的 Poisson 分布。

由 Poisson 分布的性质可知, $E[F(t)] = \lambda t$,认为 λ 是单位时间内发生的事件的平均次数,一般称 λ 是 Poisson 过程的强度或速率,即“发生率”。在本文中 λ 表示集群故障发生的次数。

2.2 集群损害的指数分布

定义 2 集群损害。每当 Hadoop 集群发生一次集群故障即认为集群系统受到了一定程度的集群

损害。

在时间一定条件下集群故障发生次数 $\{F(t), t \geq 0\}$ 符合 Poisson 过程,如图 1 所示的一条样本路径为跳跃度为 1 个单位量损害均值的阶梯函数。 T_n 表示第 n 次 ($n = 1, 2, \dots$) 集群故障发生的时刻,并规定 $T_0 = 0$ 。

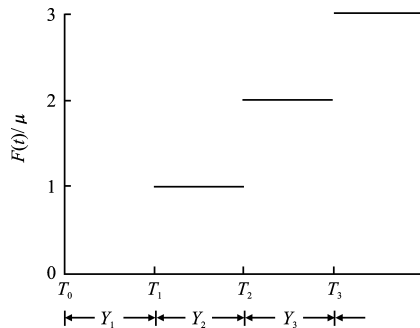


图 1 集群故障 $F(t)$ 的样本路径

令 Y_n 表示 T_n 与 T_{n-1} 时间间隔内发生的集群故障造成的集群损害。因此, Y_n ($n = 1, 2, 3, \dots$) 中 Y 服从参数为 λ 的指数分布,且相互独立。当指数分布函数中 $n = 1$ 时是 Γ 分布的一种特殊形式, Γ 分布公式为

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt.$$

其计算可满足性质 $\Gamma(n+1) = n\Gamma(n)$,是阶乘在实数上的泛化,由 Γ 分布的独立可加性,易得 Y_n 服从参数为 n 和 λ 的 Γ 分布。

计数过程 $\{F(t), t \geq 0\}$ 是参数为 λ 的 Poisson 过程,每次时间间隔发生的集群故障造成的集群损害 $Y_1, Y_2, \dots$ 相互独立,且该过程服从同一参数为 λ 的指数分布^[6]。由此可知,对计数过程 $\{F(t), t \geq 0\}$ 进行计算机模拟计算的方便途径:只需要选取合适的量化后的集群损害值作为指数分布的参数,并将指数分布的计算结果作为 Y_i ($i = 1, 2, 3, \dots$),即可得到 Poisson 过程的一条样本路径。

3 Hadoop 集群平均寿命

3.1 集群平均寿命计算方法

定义 3 集群平均寿命。当 $[0, t]$ 时间内 Hadoop 集群受到的冲击次数 $F(t)$,形成参数为 λ 的 Poisson 过程。每次冲击造成的损害 Y_i ($i = 1, 2, 3, \dots$) 独立同指数分布,均值为 μ 。那么到某时刻 t 为止系统受到的总伤害为

$$X(t) = \sum_{i=1}^{F(t)} Y_i.$$

因为损害会累积,当总损害超过一定极限 A 时,

Hadoop 集群将终止运行. 假设到时刻 T 为止系统受到的总伤害达到极限 A , 可求得

$$\begin{aligned} P\left\{\sum_{i=1}^{F(t)} Y_i \leqslant A\right\} &= P\left\{\sum_{i=1}^n Y_i \leqslant A\right\} \\ &= \int_0^A \frac{\mu^{-n}}{\Gamma(n)} x^{n-1} e^{-\mu^{-1} x} \mathrm{d} x, \end{aligned}$$

则集群的平均寿命为

$$E(T)=\int_0^{\infty} P\left\{\sum_{i=1}^{F(t)} Y_i \leqslant A\right\} \mathrm{d} t .$$

求解上式并简化得到预测的 Hadoop 集群平均寿命 $E(T)$:

$$E(T)=\frac{\mu+A}{\lambda \mu} .$$

3.2 集群平均寿命影响因素

由 3.1 节计算出的集群平均寿命结果可知, 若集群故障单位量损害均值 μ 和损害极限 A 一定时, 集群受到冲击次数越多, 则集群故障的次数越多, 集群寿命也越低. 而集群故障损害的均值大小与 Hadoop 集群所执行的任务类型和数据量大小密切相关, 即在相同环境的 Hadoop 集群下运行不同任务类型和不同数据量大小的测试程序得到的集群故障损害的均值是不同的.

在 Hadoop 集群下执行的任务类型主要包括以下几类: I/O 密集型 (I/O-bound)、CPU 密集型 (CPU-bound) 和综合型 (Synthesizing 类型). 而 Hadoop 集群可以处理的数据量大小可根据其自身环境情况自行设置.

4 实验过程与分析

4.1 实验环境

本文实验环境操作系统版本为 Ubuntu12.10, 并且安装 JDK1.7.0_02 和 Eclipse SDK3.7.0 集成开发环境, 搭建软件版本为 hadoop-0.20.2 开源 Hadoop 集群, 且所有实验均采用统一软件版本进行测试. 集群的结构均为 16 节点 Hadoop 集群: 1 台 Master 节点 (hadoop01), 其余 15 台作为 Slave 节点 (hadoop02, hadoop03, hadoop04... ..).

在此基础上, 分别配置出三种不同特征类型的 Hadoop 集群作为分析集群平均寿命的实验对比平台. 三种集群分别为默认配置的开源 Hadoop 集群; 自适应任务调度方法^[7]优化的 Hadoop 集群, 该方法是根据节点性能情况计算节点权值比例参数对节点进行任务量分配; 引入 POH 模型方法^[8]的 Hadoop 集群, 该模型通过利用 Markov 随机过程的预估思想对集群未来性能做出预测, 并在集群性能迅

速下降之前及时对集群进行调整而达到优化效果.

4.2 集群故障参数采集

为了使数据信息更加全面准确, 本文实验分别采用 hadoop-2.20.2-examples.jar 中典型的三种典型测试程序: Grep-Count (CPU-bound)、TeraSort (I/O-bound) 和 WordCount (Synthesizing). 并选取实验数据量分别为 4 GB、8 GB、12 GB、16 GB 和 20 GB 情况, 在时间相同的条件下在三种不同 Hadoop 集群实验对比平台连续执行任务, 并从多组实验结果中求取平均值作为统计结果. 表 1 中记录了不同测试程序、实验数据量以及对比平台下集群故障的次数, 其中“/”表示该类测试过程中没有出现集群故障情况.

表 1 集群故障信息统计表

测试程序	数据量/ 节点	Open Hadoop	自适应 任务调 度方法	POH 模型
TeraSort (I/O-Bound)	4 GB	2	/	/
	8 GB	4	2	1
	12 GB	6	3	2
	16 GB	9	5	3
	20 GB	12	6	5
Grep-Count (CPU-Bound)	4 GB	1	/	/
	8 GB	3	1	/
	12 GB	5	2	1
	16 GB	6	3	3
	20 GB	8	4	4
WordCount (Synthesizing)	4 GB	2	/	/
	8 GB	4	2	1
	12 GB	6	3	2
	16 GB	8	4	3
	20 GB	10	5	4

由表 1 可知, z 开源 Hadoop 集群在三种任务型的程序测试过程中均会较频繁地出现故障, 尤其是在 I/O-Bound 型的测试中, 当任务执行时间相同情况下, 数据量为 20 GB, 平均集群任务执行总次数为 40 次时, 故障数达到了总执行次数的 30% 还多. 而基于节点能力的任务级调度自适应分配方法优化后的 Hadoop 集群在程序测试过程中平均集群任务执行总次数为 41 次, 集群故障次数出现次数较少, 最高故障数仅为总执行次数的 14.6%, 而引入 POH 模型优化后的 Hadoop 集群平均集群任务执行次数为 42 次, 故障次数最少, 仅占总执行次数的 11.9%.

4.3 实验结果及分析

由集群寿命的计算方法, 以选取比较典型的任务数据量为 20 GB 的 Hadoop 集群故障信息统计情

况为例,利用 Poisson 过程计算并分析得到三种不同 Hadoop 集群在不同类型测试程序下的集群平均寿命结果对比,分别如图 2、图 3 和图 4 所示。

(1)I/O 密集型

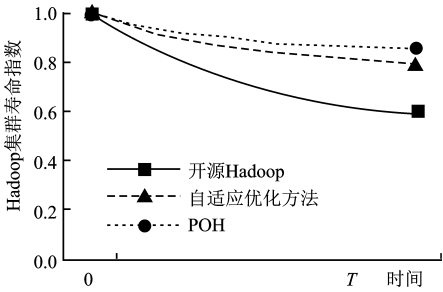


图 2 TeraSort 程序测试 Hadoop 集群寿命对比

(2)CPU 密集型

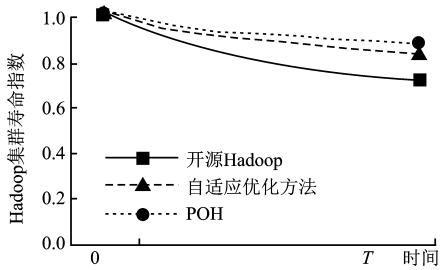


图 3 Grep-Count 程序测试 Hadoop 集群寿命对比

(3)综合型

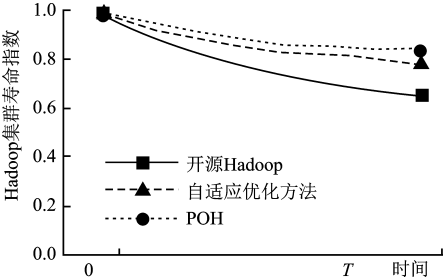


图 4 WordCount 程序测试 Hadoop 集群寿命对比

由多组对比实验结果可知,通过 Poisson 过程预测得到的 Hadoop 集群平均寿命是比较准确的,且随着集群故障发生次数的增加,集群平均寿命会呈现不同幅度的减少.在相同的 Hadoop 集群实验平台下,集群平均寿命减少的趋势与 Hadoop 集群执行的任务类型有关;若在不同的 Hadoop 集群实验平台下,集群平均寿命则与集群的性能密切相关.显然,具有良好性能的 Hadoop 集群能够拥有较为持久的集群平均寿命,而性能较差的 Hadoop 集群在长时间使用但不经适当的优化与维护的情况下,集群平均寿命下降趋势明显。

因此,利用 Poisson 过程预测 Hadoop 集群平

均寿命为集群性能进行进一步优化与提高奠定了良好的基础。

5 结束语

本文通过分析研究发现 Hadoop 集群故障的发生和造成集群损害的过程符合 Poisson 分布,合理地利用 Poisson 过程的性质,提出了一种 Hadoop 集群平均寿命预测计算方法.随着 Hadoop 集群优化方法研究的不断进步与深入,不同类型的 Hadoop 集群性能具有较大的提升,因此,将该方法作为集群性能评价标准并对于 Hadoop 集群进行深入优化具有重要的意义。

Hadoop 集群平均寿命的影响因素还有很多,对于 Hadoop 集群平均寿命的预测需要考虑的角度还应该更加综合与全面,在未来的工作中会进行更深入的分析与扩展。

参考文献:

[1] 王淑芬,高军礼,邹普,等. 基于 Hadoop 的广域网分布式主题爬虫系统框架[J]. 计算机工程与科学,2015, 37(4):670-675.

[2] 应毅,任凯,刘正涛. 基于云计算技术的数据挖掘[J]. 微电子学与计算机,2013,30(2):161-164.

[3] P Szul, T Bednarz. Productivity frameworks in big data image processing computations - creating photographic mosaics with Hadoop and Scalding [C]// Proceeding of the 14th International Conference on Computation Science, Elsevier;IEEE,2014: 2306-2314.

[4] Lorenzo Rossi, Jacob Chakareski, Pascal Frossard, Stefania Colonnese. A Poisson Hidden Markov Model for Multiview Video Traffic [J]. IEEE/ACM Transactions On Net Working,2014,23(2):1063-6692

[5] Jian Sun, Hongyan Zhu, Zongben Xu, et al. Poisson image fusion based on Markov random field fusion model [J]. Information Fusion,2013,26(14):241-254.

[6] 张波,商豪. 应用随机过程(第三版)[M]. 北京:中国人民大学出版社,2014.

[7] 郑晓薇,项明,张大为,等. 基于节点能力的 Hadoop 集群任务级调度自适应分配方法[J]. 计算机研究与发展,2014,51(3):618-626.

[8] Xiaowei Zheng, Jiaxuan Li, Yong Zhang, et al. An optimization model of Hadoop cluster performance prediction based on Markov process [J]. International Journal of Computer Systems Science & Engineering, 2015,24(6):724-734.

硬件触发的两倍多,软件触发的滞后性不能及时地响应系统的请求,不利于整个 SoC 的正常工作.因此,利用硬件触发的这种机制不仅对于 PWM 模块有着天然的好处,而且能够进一步提高系统的实时性.

仿真结果表明该 PWM 模块的功能不仅与预期中的功能相吻合,而且在 IP 核可配性和实时性上对比有了明显的提高.在仿真平台上验证之后,还可以继续在 FPGA 板上进行板级验证.因为在板级验证过程中,近似真实的工作环境可能会把软件仿真中没有能够暴露出来的问题给显露出来.这样软硬件协同验证,更符合整个验证工作的流程,最大化地排除设计上的缺陷.

5 结束语

本文设计的 PWM 模块不仅本身就具有可靠的通用性,接口设计灵活,可移植性好,而且有着多种输出模式下的高可配性和良好的实时性;另一方面还能跟其他 IP 核进行互联通信,以提高各 IP 核之间的互联性.另外,本设计并没有占用过多的硬件资源,最大限度地减少了芯片的面积.此设计比传统的 PWM 模块有优势,但仍然存在着许多问题需要解决,如 PWM 模块输出的稳定性,优化设计算法和计数器同步问题等等.这些问题需要我们下一步去改善提高.

参考文献:

- [1] 周悦. 基于 SoC 的 IP 软核设计与验证[D]. 西安:西安电子科技大学,2007.
- [2] 宋延强,刘川来,李思昆,等. SoC 设计中的 IP 核复用技术研究[J]. 青岛化工学院学报,2003,23(3):260-263.
- [3] 包明,包奎. 实现 PWM 脉宽调制的 FPGA 芯片研制[J]. 电工技术,2003,23(6):30-31.
- [4] 曹阳,李晓辉,王力纬,等. 基于 AMBA 总线的脉冲调制控制器 IP 核的设计与实现[J]. 武汉大学学报(理学版),2007,53(3):369-274.
- [5] 吴胜龙. 基于 APB 总线的通用 PWM 核设计与实现[D]. 济南:山东大学,2010.
- [6] 胡海兵,姚文熙,江辉鸿,等. 任意路通用 PWM 波形发生器的设计[J]. 电工技术学报,2008,23(12):115-119.
- [7] 陈伟,邹云屏,耿攀. 三相 PWM 整流器的死区效应分析及补偿方法[J]. 高电压技术,2009,35(5):1156-1163.
- [8] 邓清秀,文小玲. 基于 FPGA 的三相 PWM 发生器的实现方法探讨[J]. 武汉化工学院学报,2006,28(1):67-70.
- [9] Miguel A Sanchez, Marisa Lopez-Vallejo, Carlos A iglesias, et al. Improving hardware reuse through XML-based interface encapsulation[C]// 2012 IEEE 17th International Conference on Engineering of Complex Computer Systems. Paris:IEEE,2012:49-56.
- [10] 史江义. 基于 IP 核的 SOC 设计关键技术研究[D]. 西安:西安电子科技大学,2007.
- [11] 候立刚. SOC 芯片低功耗设计及功耗估计若干问题研究[D]. 北京:北京工业大学,2007.

作者简介:

翁崇杰 男,(1990-),硕士研究生. 研究方向为数字集成电路设计.

熊晓明(通讯作者) 男,(1959-),教授. 研究方向为集成电路设计及电子设计自动化.

E-mail: xmxiong@gdut. edu. cn.

马 德 男,(1985-),讲师. 研究方向为片上系统设计.

(上接第 153 页)

作者简介:

李佳宣 女,(1990-),硕士研究生. 研究方向为并行计算、云计算、分布式集群系统. E-mail: lijiaxuan09@aliyun. com.

郑晓薇 女,(1957-),教授. 研究方向为并行计算、云计算、分布式集群系统.

张 永 男,(1975-),副教授. 研究方向为云计算、数据挖掘.