文章编号:1671-4598(2018)12-0150-04 DOI:10.16526/j.cnki.11—4762/tp.2018.12.034 中图分类号:TP96 文献标识码:A

基于热点数据块的动态副本调整策略

吕海燕,周立军,赵 媛,张 杰
(海军航空大学 航空基础学院,山东 烟台 264001)

摘要:副本管理策略对于分布式存储系统的可用性、可靠性和系统整体性能有至关重要的作用;针对基于文件的动态副本调整策略的不足,提出了一种基于热点数据块的动态副本调整策略;根据时间局部性原理和数据访问规律,通过对历史访问周期和当前周期赋予不同的权重,数据块下一周期的预测进行访问频率计算,接下来基于计算出的预测访问频率对数据块进行热点判定;结合HDFS中数据访问规律近似二八定律的特点和热点数据块的判定结果,来确定数据块的调整阈值;最后,分3个步骤对基于热点数据块的动态副本调整策略进行性详细设计;实验结果表明,提出的基于热点数据块的动态副本调整策略在数据访问效率和集群存储资源利用率两方面有了明显提升。

关键词:访问频率;副本调整阀值;热点数据块;动态调整策略;数据访问效率;集群存储资源利用率

Dynamically Adjusting Strategy of Replica Based on Hot Data Block

Lv Haiyan, Zhou Lijun, Zhao Yuan, Zhang Jie
(Aviation Foundation Department, Naval Aviation University, Yantai 264000, China)

Abstract: Replica management strategy is critical for the availability, reliability, and overall performance of the distributed storage system. Aimed at the shortage of the dynamic replication strategy based on heat files, a strategy of dynamically adjusting replica factor base on hot data block is proposed. According to the principle of temporal locality and the rule of data access, calculate the predicted access frequency of data block by assigning different weights to the historical visit period and the current period. Then, according to the fact that HDFS data access approximately conform to twenty—eight law to determine the hot data block decision threshold, thereby determining whether a single data block is hot, dynamically adjusting the replication factor of the data block based on the determination result. Finally, the dynamic replica adjustment strategy of hotspot data block is designed in three steps. Experimental results show that the proposed dynamic replica adjustment strategy based on hot data block has improved both data access efficiency and cluster storage resource utilization.

Keywords: access frequency; replica adjusting threshold; hot data block; dynamically adjusting strategy; data access efficiency; cluster storage resource utilization

0 引言

资料显示,针对分布式存储系统的数据存储策略的研究,国内外研究者主要从数据可靠性、用户访问效率、存储资源利用率及系统负载均衡4个方面进行。分布式存储主要涉及到如何确定数据副本系数和如何在集群系统中选择数据副本存放的节点两个方面的内容。文献[1-9]从不同的方面给出了多个动态副本改进策略,这些改进策略在副本管理的某些方面都有一定的效率提升,并且主要都是依据数据的历史访问情况来预测数据访问热度,进而动态调整数据副本的相应系数。但是,这些动态调整策略大部分都是基于文件进行动态副本管理的,这样就忽略了一种事实,即很多时候用户感兴趣的不是整个文件,而只是文件中的一部分数据,即其中的一个数据块或多个数据块。因此,若是仅仅以文件为单位进行动态副本调整,其效果并不理想,不仅会导致集群系统存储资源利用率的降低,同时还会因为副本维护一致性增加而增加系统的系统开销[10]。

1 基于文件的动态副本调整策略的不足

基于文件的动态副本调整策略主要是通过对文件的历史访问情况进行分析,在此基础上对文件的访问趋势进行预测,进而动态调整其相应的副本系数。对于热点数据,通过合理地增加其副本系数来提高其访问并行度;而对于访问频率较低的数据,在保证其可靠性的前提下通适当地减少其副本数,从而降低了维护其数据副本一致性带来的系统开销,进而提高系统资源利用率[11]。事实情况,也是如此,在大多数情况下用户感性趣的只是某个文件中的某一部分数据,而非整个数据文件。因此,如果以文件为单位进行副本系数动态调整,会带来以下弊端:

1)集群系统存储资源利用率降低。HDFS分布式文件系统中主要存储的是大数据文件,仅仅因为某个大文件中的部分数据块访问频率较高,就将该整个文件被判断为热点文件,从而增加该其副本系数,这将大大浪费集群存储资源,降低系统资源利用率。

2)数据副本一致性维护成本增加。数据副本的一致性

收稿日期:2018-03-24; **修回日期**:2018-05-22。
基金项目:2016海军院校和训练机构教学成果立项项目,中国人民解放军海军参谋部,发文号参训[2016]33号。
作者简介:吕海燕(1983-),女,山东淄博人,硕士,副教授,主要从事计算机教学和计算机软件技术方向的研究。

是保证用户能够正确访问到所请求的数据的重要因素。随着热点文件数据副本数量的增加，势必会使副本管理难度增加，尤其是系统在维护数据副本一致性时所产生的消耗也会相应增加。

因此，本文提出了一种基于热点数据块的动态副本调整策略，在一定程度上克服了基于文件的副本动态调整策略的弊端，保证数据可靠性的同时进一步提升集群系统存储资源的利用率，减少数据副本一致性维护的成本。

2　基于访问率的热点数据块的判定

对于数据文件的访问情况而言，某个数据块的以往周期的访问次数有可能在某个周期某个时间段内相对较高，但在此之后不断下降，而其总体访问次数相对于其他数据块可能还是处于高水平。因此，对热点数据块的判定，不能仅仅依据其在某个时间段的历史访问次数或者平均访问次数。同时，根据时间局部性原理和数据访问规律可知，一般情况下，数据访问规律在一段时间内是今本是保持不变的，且最近被频繁访问的数据很有可能在接下来的一段时间内还会被访问到[12]。基于这种数据访问的局部性原理，本文提出了一种根据基于热点数据块的动态调整策略，通过对数据块前 N 个历史周期的访问情况赋予不同的权重合理地计算机数据块的预测访问频率，进而更加准确地对热点数据块做出判定。

2.1　计算数据块访问频率

在名字节点（NameNode）上设置一个记录数据块访问情况的三元组，主要包含：数据块标识（BID），上一周期访问频率（Pre_F），本周期访问次数（N）。由于 HDFS 集群中存储着海量的数据文件，如果对每个数据块的所有历史周期访问情况都一一记录下来，这将给 NameNode 节点带来巨大的负担。因此，为减轻 NameNode 节点的负载，这里只记录数据块上一周期的访问频率和当前周期的访问次数。接下来，通过设置平衡因子，对历史访问周期和当前访问周期赋予不同的权重，进而计算出下一周期数据块的访问频率。如果，数据块下一周期预测频率 B_f 表示，数据块上一周期访问频率用 B_f_pre 表示，数据块当前周期内的访问次数用 M 表示，访问周期用 T 表示，平衡因子用 τ。则数据预测访问频率的 B_f 计算方法如公式 1 所示。

$$B\_f = \tau B\_f_{pre} + (1-\tau)\frac{M}{T} \quad 0 \leqslant \tau \leqslant 1 \qquad (1)$$

其中：τ 的取值范围为 0～1 之间，其作用是来平衡当前访问周期和历史访问周期对预测频率影响的权重。在设置权重时，遵循上述数据访问的局部性原理，其值得设置要使得越是靠近当前周期的历史访问周期的访问频率对预测访问频率影响越大；反之，越是久远的历史周期的访问情况对预测访问频率影响影响越小。即 τ 的取值越接近 0，表示下一周期的访问频率预测更取决于当前周期内该数据块的访问情况；反之，τ 的值越接近于 1，则表示对数据块下一周期的频率预测更侧重于其过往周期访问情况即历史访问率。需要根据实际情况需求，合理地设置的值，以便准确的预测数据块下一周期的访问情况，从而确定其是否为热点数据块，进而合理地动态调整其副本系数。

根据公式（1），可以得到数据块第 i 个周期的访问频率 B_F_i 计算公式如下：

$$B\_F_i = \tau B\_f_{i-1} + (1-\tau)\frac{M_i}{T} \quad 0 \leqslant \tau \leqslant 1 \qquad (2)$$

第 $i-1$ 个周期的访问频率 B_F_i 计算公式如下：

$$B\_F_{i-1} = \tau B\_f_{i-2} + (1-\tau)\frac{M_{i-1}}{T} \quad 0 \leqslant \tau \leqslant 1 \qquad (3)$$

将公式（3）代入公式（2）可得：

$$B\_F_k = \tau^k B\_f_0 + \frac{(1-\tau)}{T}\sum_{i=1}^{k} * M \quad 0 \leqslant \tau \leqslant 1 \qquad (4)$$

其中：B_F_0 表示数据块被创建时的访问频率，因为数据块刚被创建时，前期没有历史访问情况，所以其值可置为 0。因此，数据块预测频率的计算公式（4）可简化为：

$$B\_F_k = \frac{(1-\tau)}{T}\sum_{i=1}^{k}\tau^{k-1} * M_i \quad 0 \leqslant \tau \leqslant 1 \qquad (5)$$

如式（5）所示，i 越大表示该访问周期离当前周期越，近由于介于 0 和 1 之间，所以 τ^{k-1} 的值就越大，因此，其对数据块预测访问频率的影响也就越大；反之，i 越小，表示该访问周期离当前周期越远，所以 τ^{k-1} 的值就越小，因此，其对数据块访问频率的影响也就越小。这正与前面所述的局部访问性原理相符。

2.2　判定数据块类型

研究发现，HDFS 集群系统中用户对数据的访问规律接近于二八定律，即 20％的数据占有 80％的用户数据访问量[13]。依据此定律和上述计算出的数据块访问频率，将访问率前 20％的数据块判定为热点数据块，对其副本数量进行相应的增加，以提高用户访问效率；将访问率后 20％的数据块判定为冷门数据块，在保证其数据可用性的前提下对其副本数量进行适当地减少，以节省系统开销；将剩余的数据即访问率在中间 60％的判定为正常数据块，其副本数量在当前周期内保持不变，随着访问周期的推进，正常数据块有可能会在将来某个访问周期成为热点数据或冷门数据。

表 1　数据块类型判定表

数据块类型	数据块访问频率/％	副本调整策略
热点数据块	80％≤B_f≤10％	增加
正常数据块	20％≤B_f≤80	不便
冷门数据块	0％≤B_f≤20	减少

3　数据块副本动态调整系数的确定

数据块副本系数的动态调整范围不能太大也不能过小。如果副本系数过大，虽然可以在一定程度上提升数据访问的可用性和并行度，但同时也会大大增加系统副本维护消耗，从而降低存储资源的利用率；如果副本系数过小，则无法保证数据的可用性和可靠性[14-16]。因此，合理地确定

副本系数的动态调整的范围或者阀值是关键。

3.1 数据块副本访问频率

如前所述，每个数据块的访问频率可通过（2）计算出，并且计算出的这个访问频率是整个数据块的访问频率。由于在 HDFS 中，数据块是以多副本的方式存储的，并且用户对某个数据块的访问请求将尽可能地均匀分布在该数据块的所有副本上。所以，假设某数据块的副本数量为 n，则该数据块每个副本的访问频率则为：

$$REP\_f = \frac{B\_f}{n} = \frac{\tau B\_f_{pre} + (1-\tau)\frac{M}{T}}{n} \quad 0 \leqslant \tau \leqslant 1 \tag{6}$$

3.2 数据块副本调整阀值

副本调整阀值可分为增加阀值（Inc_th）和减少阀值（Dec_th）。为有效节省系统计算和内存资源，依据二八定律本文采用如下方法确定副本的调整阀值。

依据（6）计算出的数据块副本的访问频率，对所有该数据块的所有副本访问频率求平均，得到 $Rep\__{favg}$ 即数据块副本平均访问频率，然后将增加阀值 Inc_th 设定为数据块副本平均访问频率的 2 倍，将减少阀值 Dec_th 设定为数据块副本平均访问频率的 1/2。假设数据块副本系数为 n，则 Inc_th 和 Dec_th 的计算公式如（7），（8）所示：

$$Inc\_th = 2 * Rep\__{favg} = \frac{2}{n}\sum_{i=1}^{n} Rep\__{fi} \tag{7}$$

$$Dec\_th = \frac{1}{2} * Rep\__{favg} = \frac{1}{2n}\sum_{i=1}^{n} Rep\__{fi} \tag{8}$$

采用此方法确定的副本调整阀值不能完全符合二八定律，但是由于 HDFS 集群系统中数据的访问规律也只是近似符合二八定律，并且该计算方法能够迅速得到结果，因此，本文采用了该方法来确定副本调整阀值。

4 基于热点数据块的动态副本调整策略

对于热点数据块，依据计算出的数据块增加阀值（Inc_th）增加其副本数量，以提高用户访问量，均衡数据节点 $DataNode$ 的负载；对于冷门数据块则依据计算出的数据块减少阀值（Dec_th）减少其副本数量，在保证数据可用性的前提下，降低副本一致性维护消耗、节省系统存储开销。本文设计的基于热点数据块的数据块动态副本调整策略如图 1 所示。

步骤 1：根据各数据块访问频率计算出其相应的数据块副本增加阀值（Inc_th）和减少阀值（Dec_th）。

步骤 2：将数据块副本访问频率 Rep_f 与其副本增加阀值（Inc_th）进行比较，如果 $Rep\_f > Inc\_th$，则表示该数据块副本访问负载过重，可能会导致其所在的数据节点 $DataNode$ 负载过重，从而造成性能瓶颈。因此，需要增加其副本数量，假设需要的副本数量为 R，则应满足：

$$\frac{B\_f}{R} < Inc\_th \tag{9}$$

同时 R 的取值不能超过系统的最大副本系数。根据计算出的副本数量，创建新的数据块副本，并依据相应的副本放置策略选择最佳的数据节点 $DataNode$ 放置新的副本。

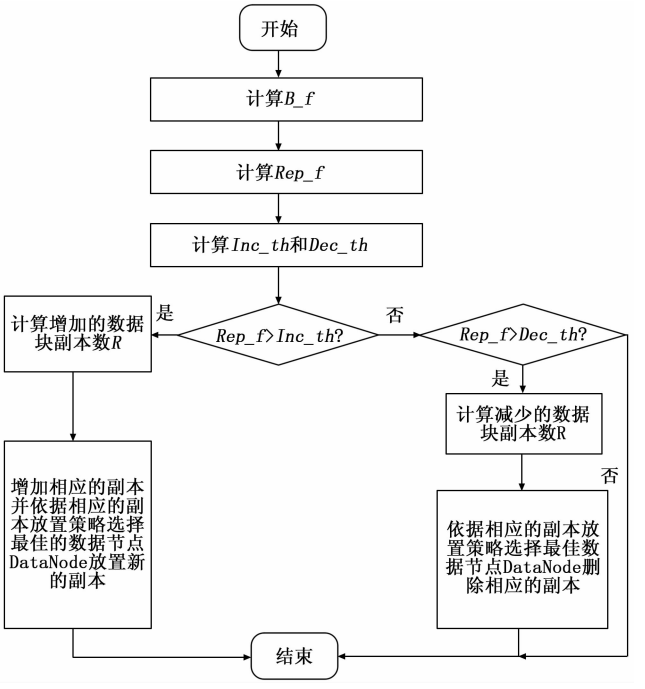

图 1 基于热点数据块的动态副本调整策略流程图

步骤 3：将数据块副本访问频率 Rep_f 与其副本减少阀值（Dec_th）进行比较，如果 $Rep\_f < Dec\_th$，则表示该数据块副本访问负载过少，为降低副本一致性维护消耗、节省系统存储开销，在保证数据可用性的前提下，需要减少其副本数量，假设需要的副本数量为 R，则应满足：

$$\frac{B\_f}{R} > Dec\_th \tag{10}$$

同时 R 的取值不能小于系统的最小副本系数。同样，根据计算出的副本数量，依据相应的副本放置策略选择最佳的数据节点 $DataNode$ 删除副本。

5 实验验证与结果分析

由于本文主要针对数据访问效率和集群存储资源利用率两方面的改进提出了基于热点数据块的动态副本调整策略，所以改进策略效果的验证实验的设计主要从上述两个方面进行。并将实验结果与 HDFS 固定副本策略和基于文件的动态副本调整策略进行比较分析，以验证本文设计的基于热点数据块的动态副本调整策略的有效性。

5.1 小型集群实验环境设计

选取实验室 8 台机器部署 HDFS，机器配置如表 2 所示。其中，192.168.2.253 作为 NameNode，其余 7 台作为 DataNode，并依据 IP 将其分布在两个 rack（机架）中，机架内部使用百兆以太网。

5.2 实验结果

采用默认副本系数 3，对随机生成的 100 个文件中选取其中一个大小为 256M（共 4 块）的文件，对其进行 7 个周期的访问测试，各数据块各周期访问频率情况如表 4 所示。

表 2　小型集群环境计算机配置

机架	IP	角色	CPU	内存	硬盘
Rack1	192.168.7.10	NameNode	intel i7 四核	8G	1T
	192.168.2.23	DataNode	intel i5 双核	4G	500G
	192.168.2.24	DataNode	intel i5 双核	4G	500G
	192.168.2.25	DataNode	intel i5 双核	4G	500G
Rack2	192.168.2.65	DataNode	intel i5 双核	8G	1T
	192.168.2.66	DataNode	intel i5 双核	8G	1T
	192.168.2.67	DataNode	intel i5 双核	8G	1T
	192.168.2.68	DataNode	intel i5 双核	8G	1T

表 3　数据准备及相关参数设置

数据文件	平衡因子	数据块副本数量上限	数据块副本数量下限	默认副本系数	访问周期
随机生成 100 个平均大小 192M 块大小 64M	$\alpha=5$	$R_{max}=2$	R_{min}	$R=3$	$T=2$ 小时

表 4　大小为 256M 的文件各数据块 5 个周期访问频率表

周期 \ 数据块	A	B	C	D
1	0.42	0.47	0.21	0.58
2	0.41	0.76	0.23	0.52
3	0.38	0.92	0.22	0.54
4	0.43	0.94	0.19	0.51
5	0.41	0.84	0.20	0.53
6	0.40	0.27	0.23	0.49
7	0.37	0.38	0.19	0.52

由表 4 可以看出，数据块 B 为热点数据块。该访问情况下 3 种副本策略文件占用集群存储空间情况如图 2 所示，可以看出，不同的副本管理策略下，随着访问周期的不同文件所占集群存储空间的变化是不同的。在固定副本管理策略中，文件所占的存储空间固定不变；基于文件的副本调整策略使得当文件中的某个数据块被频繁访问时，整个文件都将成为热点文件，其所有数据块的副本均有所增加，导致文件所占的集群存储空间持续增加，只有当文件的所有数据块均被判定为冷门数据块时，系统文件默认副本系数才会变化，如在第 6 个访问周期，副本系数变为 2，整个文件存储空间变为最小；而本文提出的基于热点数据块的动态副本调整策略，不会因为数据块 B 的高访问率，导致整个文件占用的存储空间增加，对于热点数据块 B，只增加该数据块的副本数量，不会影响该文件其它数据块的副本数量变化。实验结果证明，本文提出的基于热点数据块的动态副本调整策略在集群存储空间利用率方面优于基于文件的副本调整策略和固定副本调整策略。

对于数据访问效率，以数据块 B 为主要测试对象，随着不同周期访问频率的变化，3 种策略下其平均响应时间情况如图 3 所示。可以看出，在第 1 个周期，由于数据块访问

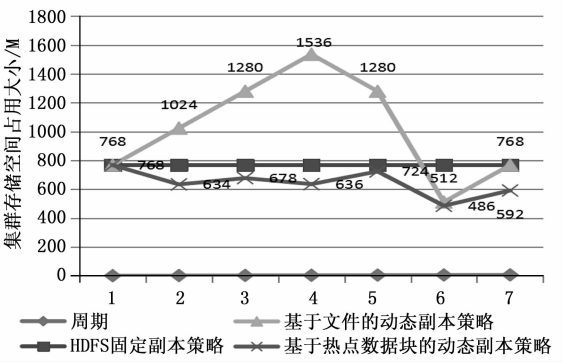

图 2　3 种策略存储空间占用大小比较图

频率较低，还没有达副本增加的条件，因此，基于文件和基于本文提出的热点数据块的动态策略的响应时间和原 HDFS 固定副本策略的平均响应时间不但没有减少，反而略高。这是由于动态副本调整策略每个周期都会对文件或数据块进行访问频率的计算，从而进行相应的热点判断，这会耗费一部分时间。之后随着访问频率的增加，3 种策略的响应时间均会达到一个最大值，这主要是由于随着访问频率的不断增加，会开始进行副本增加操作，这同样会占用一部分系统资源。接下来，增加后的副本开始分担相应的数据访问请求，因此，动态副本调整策略的响应时间的优势凸现出来，而本文提出的基于热点数据块的响应效率优势更加明显。

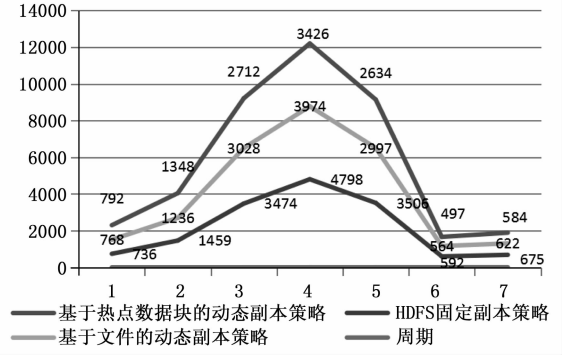

图 3　3 种策略数据访问效率比较图

6　小结

本文在总结分析基于文件的动态副本调整策略不足的基础上，提出了一种基于热点数据块的动态副本调整策略。该策略在时间局部性原理前提下，通过对数据块的历史访问周期的访问率赋予不同的权重进而预测出数据块下一周期的访问频率。然后根据 HDFS 中数据访问规律近似二八定律这一特点，依据计算出的数据块预测访问频率，对数据块进行热点和非热点判定，从而将数据块分为热点数据块、正常数据块和冷门数据块三类。对于热点数据块通过增加该数据块的副本系数以增加访问并行度提高用户访问效率，同时在保证最低可靠性前提下对非热点数据块的副本

（下转第 157 页）