

基于内容分发的数据网格副本创建策略^{*}

李 飞¹, 郭昌松¹, 赵伯听², 曾 辉¹

(1. 成都信息工程学院 网络工程学院, 成都 610225; 2. 福建省气象局 气象服务中心, 福州 350001)

摘 要: 结合混合的网格拓扑模型, 提出了一种基于内容分发的副本快速创建策略。该策略使用区域中心节点之间的历史记录和传输代价, 选择一个最佳的区域中心, 将新产生的文件传输到最佳区域中心; 在客户节点从其他的区域中心传输某一文件的时候, 同时在该客户节点所在的区域中心建立文件的副本。在网格模拟器 OptorSim 中测试了该策略, 结果表明该策略能够减少作业的平均完成时间。

关键词: 内容分发; 副本创建; 传输代价

中图分类号: TP399 **文献标志码:** A **文章编号:** 1001-3695(2012)10-3908-03

doi: 10.3969/j.issn.1001-3695.2012.10.081

Strategy of data grid replica creation based on content distribution

LI Fei¹, GUO Chang-song¹, ZHAO Bo-ting², ZENG Hui¹

(1. Dept. of Network Engineering, Chengdu University of Information Technology, Chengdu 610225, China; 2. Weather Service Center, Fujian Meteorological Bureau, Fuzhou 350001, China)

Abstract: This paper proposed a strategy to quickly create a replica based on content-distribution, which used the history records between the regional central nodes and chose a best regional node where to transfer new files. When a client node received some file from other regional node, it would create the copy of the file on the regional node of this client node at the same time. It tested the strategy in the grid simulator OptorSim, which can reduce the average completion time of job.

Key words: content distribution; replica creation; transmission cost

数据网格^[1]的关键技术之一是数据副本管理, 主要涉及到副本的创建、选择、定位和一致性。副本创建是副本管理的基础, 其主要目标是通过合理地创建数据副本, 将副本尽量放在距离访问节点较近的数据节点, 降低用户访问数据的延迟, 并减少数据网格中占用的广域网带宽, 最终提高数据密集型任务的吞吐率。副本创建分为动态和静态创建, 静态创建用于副本的初始化部署, 而动态创建是根据用户的请求创建副本。副本创建要解决下面的三个问题: 什么时候创建数据副本、创建哪些数据的数据副本、在哪些节点创建数据副本。

1 相关研究工作

文献[2]提出了六种副本创建策略: no replication、best client、cascading、plain caching、caching + cascading、fast spread。实验结果表明, 除 no replication 外的五种策略都可以节省网络带宽和降低用户访问延迟, 其中 fast spread、cascading 效果相对较好, 但 fast spread 的副本占用空间较高。这六种副本创建策略考都只是简单地考虑了用户请求文件的次数, 没有考虑网络带宽和文件大小对用户访问时延的影响。文献[3]中提出了 SBU (simple bottom upper) 和 ABU (aggregate bottom upper) 两种副本创建策略, 也是基于层状网格模型, 根据用户访问的历史记录, 决定是否创建副本。当请求的次数超过预设值的阈值, 就在请求节点到原始数据节点的中间节点上, 建立数据的副

本, 可见该算法结合了 fast spread 策略。文献[4]提出了基于经济模型的副本创建策略, 按照反向拍卖协议确定副本创建位置并进行副本选择, 它将文件传输时间作为拍卖的价格指标。该策略针对欧洲数据网格设计, 并没有考虑各节点的存储能力, 在节点存储能力较小时容易引起副本的频繁建立与删除。文献[5]提出了一种基于存储联盟的副本创建策略, 该策略以存储联盟为基本单位进行数据存储, 该策略适用于网络带宽比较小并且单个节点存储能力比较弱的数据网格。文献[6]提出了基于 hybrid 拓扑的数据网格副本创建策略: 增强的最小化时间算法, 该策略考虑了网络带宽、网络传输延迟和文件的请求次数。此后的研究提出了基于副本效益和副本代价的副本创建策略^[7]、基于存储域的副本创建策略^[8]、启发式的副本放置策略^[9]等等。

通过以上分析可知, 副本的创建策略与数据网格的拓扑结构、原始数据存放的位置、文件的请求次数、文件大小、节点之间的网络带宽、节点的存储能力等因素有密切关系。本文提出了一种基于内容分发的副本快速创建策略, 该策略结合了动态创建和静态创建的优点, 考虑了数据传输带宽、用户请求次数, 减小了数据传输时延。

2 数据网格的拓扑模型

由于数据副本的创建与网格的拓扑结构有密切的关系, 受

收稿日期: 2012-02-13; **修回日期:** 2012-04-05 **基金项目:** 四川省科技支撑项目(2011GZ0195)

作者简介: 李飞(1966-), 男, 湖南常德人, 教授, 硕士, 主要研究方向为云计算、信息安全; 郭昌松(1985-), 男, 山东德州人, 硕士, 主要研究方向为网格计算(guochangsong@126.com); 赵伯听(1984-), 男, 浙江温州人, 硕士, 主要研究方向为信息安全、数据库应用; 曾辉(1984-), 河南驻马店人, 硕士, 主要研究方向为信息安全、网格计算。

文献[5,7,8,10]的启发,提出了一种网型和树型相结合的混合式网络拓扑结构,如图 1 所示。

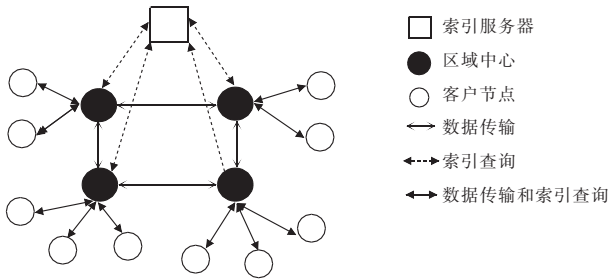


图 1 网络拓扑模型

该模型由索引服务器、区域中心和客户节点组成。索引服务器负责保存所有区域中心所有文件的索引信息。区域中心可以从索引服务器查询所要请求的文件。区域中心节点是由管理员指定的高性能计算机或者集群,具有存储空间、带宽较大的特点。客户节点和区域中心组成网格子系统,区域中心负责存储该子网格中数据副本的索引信息,即区域中心保存本地和客户节点的文件索引信息。区域中心还要负责保存本地产生的原始数据和从其他区域中心产生的数据副本。

当客户端请求某一文件的时候,先检查本地是否存在该文件,如果不存在,就向该节点所在的区域中心发送文件请求。区域中心首先查找本地的文件索引信息,如果存在该文件的索引信息,就将该请求文件的索引信息返回给客户节点;如果在区域中心也不存在该文件的索引信息,就向索引服务器请求查找该文件。索引服务器查找到该文件的索引信息,就将索引信息发送给区域中心。区域中心根据文件索引信息从其他区域中心请求文件,在该区域中心上创建文件副本,并传输数据到请求的客户节点。

3 基于内容分发的数据副本快速创建策略

基于内容分发的数据副本快速创建策略分为两个阶段:基于内容分发的副本创建阶段和数据网格副本的快速传播阶段。

3.1 基于内容分发的副本创建阶段

该阶段中副本创建是在区域中心之间进行的。每个区域中心只记录其他区域中心对该区域中心所产生文件的访问次数,形成该区域中心的历史访问记录,根据历史记录选择出前 N 个最有可能访问该区域中心数据文件的区域中心,在这 N 个区域中心中选择一个最佳区域中心。每隔一段时间 T ,区域中心就清空历史记录,并重新选择出新的最佳区域中心,将产生的新文件传输到该时间段的最佳区域中心,创建文件副本。

内容分发是指将在一段时间 T 内新产生的数据从原始数据区域中心传输到最佳区域中心。每一个区域中心都是原始数据节点,每一个区域中心都有可能成为其他区域中心的最佳区域中心。

最佳区域中心是在一段时间 T 内相对于一个原始数据区域中心而言的,在这段时间内,该原始数据区域中心将新产生的数据文件传递到最佳区域中心。最佳区域中心满足三个条件:a)最佳区域中心下面的客户节点最有可能访问从原始数据区中心传输过来的文件;b)最佳区域中心到其他有可能访问原始数据区域中心中文件的区域中心的传输代价要小;c)该区域中心有足够的存储空间。

在本文中使用区域中心之间的带宽和跳数表示传输代价,带宽越大、跳数越小,则区域中心之间的传输代价越小。在该阶段使用的基本公式包括:

$$wCost = (8 \times 2^{10}) / w \quad (1)$$

代表传输 1 GB 数据文件所花费的带宽代价。

$$y = (x - \minValue) / (\maxValue - \minValue) \quad (2)$$

归一化公式。 x, y 分别为转换前、后的值, $\maxValue, \minValue$ 分别为样本的最大值和最小值,转换后 y 为 $[0, 1]$ 之间的一个值

$$\text{cost}(i, j) = \alpha w + \beta h, \alpha + \beta = 1 \quad (3)$$

其中: α, β 分别代表区域中心 i, j 之间带宽、跳数的权重; w, h 是节点 i, j 之间的带宽和跳数各自归一化后的结果。

$$\text{transcost}(j) = \min(\text{cost}(i, j), \text{cost}(O, j)) \quad (4)$$

表示区域中心 j 到原始区域中心 O 或区域中心 i 之间传输代价的较小者。

$$E_i = \sum_{j=1}^N (\text{transcost}(j)) \quad (5)$$

表示区域中心 i 到其他区域中心传输代价的期望值。在此阶段,某一个区域中心的算法流程如图 2 所示。

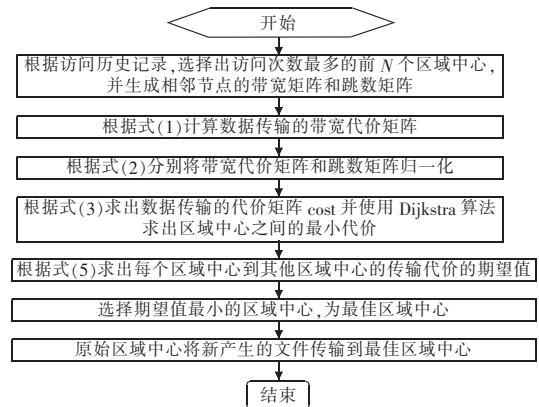


图 2 一个区域中心的流程

3.2 副本的快速创建阶段

数据网格副本的快速创建策略是客户节点主动请某一个文件,而该客户节点所在的区域中心不存在要请求的文件时,该区域中心向其索引服务器发出请求,同时保存从其他区域中心传输过来的文件,并转发给请求的客户节点。该过程就是副本的快速创建过程。

在该阶段,请求区域中心会从索引服务器获得一个逻辑文件的多个物理文件,这些物理文件分别存放在不同的区域中心,请求区域中心根据请求区域中心到其他区域中心的实时带宽,选择要传输数据的区域中心,首先检查区域中心是否有足够的空间,如果没有足够的空间,就使用 LRU 算法删除最近最少使用的缓存副本,为新的副本腾出空间。

4 仿真实验及结果分析

为分析基于内容分发数据网格副本快速创建策略的可行性和性能,本文以网格仿真模拟软件 OptorSim 为基础,使用 OptorSim 的存储单元 SE 和计算单元 CE 组成新的网格站点,在网格站点中加入一个标志,表明该站点是区域中心还是客户站点,客户站点属于哪一个区域中心站点。每个中心区域站点类里面都有一个目录管理对象,使用该对象管理本区域中心的所有文件。

在本文的网格模拟中,为了减少网格作业调度算法对副本创建策略的影响,本文中设计的每个作业只能能够在在一个客户节

点中执行,即作业中请求的文件相同,作业名称不相同。作业中文件的访问方式都来用随机访问方式,以减少文件访问方式的影响。

本文使用拥有6个区域中心和9个客户节点的网络拓扑图,具体配置如图3所示,第一列代表客户节点所属的区域中心,所有的节点从0开始编号,-1代表该节点是区域中心节点,其他数字代表该节点所属区域中心的编号;第二列表示CE的个数;第三列表示CE的类型;第四列表示SE的个数;第五列表示SE的容量,后面的几多列表示两个节点之间的带宽,0表示节点之间没有带宽。

所属的 区域中 心节点	站点的计算单 元CE和存储单 元SE的配置	两个相邻站点之间的带宽关系																			
-1	2 1.0 1	1000000	0.	100.	0.	0.	0.	200.	0.	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.
-1	2 1.0 1	1000000	100.	0.	250.	150.	0.	0.	0.	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.
-1	2 1.0 1	1000000	0.	250.	0.	150.	180.	0.	0.	0.	100.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-1	2 1.0 1	1000000	0.	150.	150.	0.	300.	0.	100.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
-1	2 1.0 1	1000000	0.	0.	180.	300.	0.	120.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	100.
-1	2 1.0 1	1000000	200.	0.	0.	0.	120.	0.	0.	0.	0.	0.	0.	0.	0.	100.	100.	0.	0.	0.	0.
3	1 1.0 1	1000000	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
3	1 1.0 1	500000	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
2	1 1.0 1	500000	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
2	1 1.0 1	500000	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
1	1 1.0 1	500000	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
0	1 1.0 1	500000	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
4	1 1.0 1	500000	0.	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
4	1 1.0 1	500000	0.	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
5	1 1.0 1	500000	0.	0.	0.	100.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.

图3 网络节点配置

在模拟实验中,假设在区域中心0新产生了10个数据文件,其余5个区域中心下面的客户节点都有可能访问新产生的数据文件。令式(3)中的 $\alpha=0.8$,表示节点之间的带宽占主要的因素。为了提高实验的准确性,避免较大的误差,实验结果取相应10次实验的平均值。

本文对比了没有副本创建策略、快速传播副本创建策略、基于内容分发的快速副本创建策略这三种副本创建策略使用作业的平均完成时间作为基本评价标准,结果如图4所示。

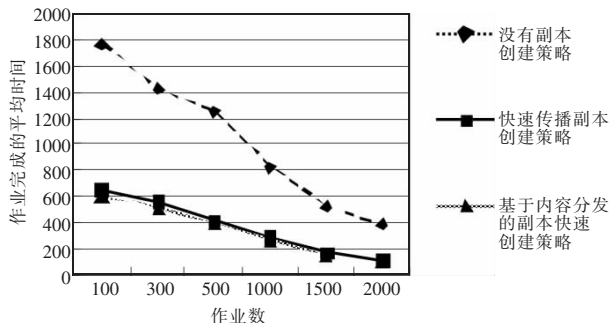


图4 三种策略的对比

从图4可以看出:快速传播的副本创建策略和基于内容分发的副本创建策略都要远远优于没有副本创建策略;在作业

重复数量比较少的时候,基于内容分发的副本快速创建策略的作业平均完成时间明显低于快速传播副本创建策略的作业平均完成时间,但是随着作业重复次数的增加,这种优势逐渐减小。

5 结束语

本文分析了多种网格副本创建策略,总结了与副本创建策略有关的因素,建立了一种新的网格拓扑模型,结合该网格拓扑模型,提出了基于内容分发的快速副本创建策略,并对该策略进行了详细描述,在OporSim模拟器中以作业的平均完成时间作为评价标准对比了没有副本创建策略、快速传播副本创建策略和基于内容分发的快速副本创建策略,实验表明基于内容分发的快速副本创建策略在副本创建方面具有一定的优势。

参考文献:

- [1] FOSTER I, KESSELMAN C. The Grid 2: blueprint for a new computing infrastructure[M]. San Francisco: Morgan Kaufmann, 2004.
- [2] RANGANATHAN K, FOSTER I. Identifying dynamic replication strategies for a high-performance data grid[C]//Proc of the 2nd International Workshop on Grid Computing. Berlin: Springer-Verlag, 2001: 80-85.
- [3] TANG M, LEE B S, YE C K, et al. Dynamic replication algorithms for the multi-tier data grid[J]. Future Generation Computer Systems, 2005, 21(5): 775-790.
- [4] CAMERON D G, CARVJAL-SCHIAFFINO R. Evaluating scheduling and replica optimization strategies in OporSim[C]//Proc of the 4th International Workshop on Grid Computing. Washington DC: IEEE Computer Society, 2003: 52-59.
- [5] 孙海燕. 数据网格副本管理关键技术研究[D]. 长沙: 国防科技大学, 2005.
- [6] 卢炎生, 胡辉. 基于 hybrid 拓扑的数据网格副本创建策略[J]. 计算机应用研究, 2007, 24(11): 286-288.
- [7] 邢长明, 刘方爱, 杨林, 等. 教育资源网格模型及副本创建策略[J]. 软件学报, 2009, 20(10): 2844-2856.
- [8] 李田来, 刘方爱, 庄波. 基于 GLOBUS 的教育网格副本创建策略研究[J]. 计算机工程与设计, 2008, 29(15): 3881-3884.
- [9] 付雄, 王汝传. 树型数据网格环境下副本放置算法研究[J]. 南京邮电大学学报: 自然科学版, 2011, 31(3): 72-78.
- [10] 姜建锦, 杨广文. 数据网格中请求呈现分组特性的副本管理策略研究[J]. 计算机研究与发展, 2009, 46(2): 186-193.
- [11] 赵秋云, 罗省贤, 魏乐. 网格环境下基于传输代价的副本优化策略[J]. 计算机应用研究, 2010, 27(4): 1444-1447.