

# 基于 Globus 平台的网格副本创建

## Globus – Based Grid Replica Location

李丽君 曹 恺 (浙江大学 计算机学院 浙江杭州 310027)

**摘要:** 为了分析不同网格副本创建策略的性能,根据基于 Globus 平台的网格副本结构,通过实验比较了三种策略的指标,体现了策略创建的多样性,为具体的访问需求提供副本创建参考。本次实验证明策略的正确与否很大程度上决定了网格的性能高低,而策略创建的多样性则表明了该项研究的任重道远。

**关键词:** 网格 副本 平台 策略 Globus

## 1 引言

### 1.1 文章安排

本文第 2 节介绍基于 Globus 平台的副本属性。第 3 节给副本创建的策略。第 4 节给出实验数据作性能对比和分析。第 5 节给出结论。第 6 节给出未来工作。

#### 1.1.1 基本介绍

网格将整个互联网整合成一台巨大的超级计算机,它实现了计算资源、存储资源、数据资源、信息资源、知识资源的全面共享。资源分布是网格最基本的特征,它通过集中分散的资源来完成计算。而正是这种集中分散的计算方式,使用户在享受资源的同时,也感受到了互联网高延时所带来的大数据访问限制。为了便于用户对分散数据的快速访问,网格系统需要根据用户的动态访问特征及时向访问端创建副本,以降低由于用户对远程数据的访问所带来的访问延时和带宽消耗。对于副本的创建,可以通过可视化的 Globus 平台来具体实现<sup>[1]</sup>。

## 2 基于 Globus 平台的副本属性

利用 Globus 平台我们可以创建出具有高可用性的网格副本,这些副本应该具有如下属性:

(1) 动态性:需要时,副本可以动态的创建和删除;

(2) 安全性:副本必须被安全地创建,以免信息泄露给第三方或是数据发生损坏;

(3) 高效性:副本必须能通过适当的时间和资源迅速创建起来;

(4) 适应性:副本处理要能够处理多种网络速度、异构存储资源、处理速度和失败;

(5) 灵活性:副本应该能够灵活地加入或者离开网格;

(6) 副本选择:在一个环境中,通常同一对象有多个副本可得,所以这时需要一个副本选择服务,主机利用它来选择一个合适的能够满足它服务需求的副本;

(7) 副本一致性:在一个环境中,副本通常需要更新,不同程度的一致性以及更新频率需要被提供;

(8) 扩展性:副本系统需要能够处理大量的副本,同时支持并发的副本创建。

其对应的 Globus 平台网格副本结构如下<sup>[2]</sup>:

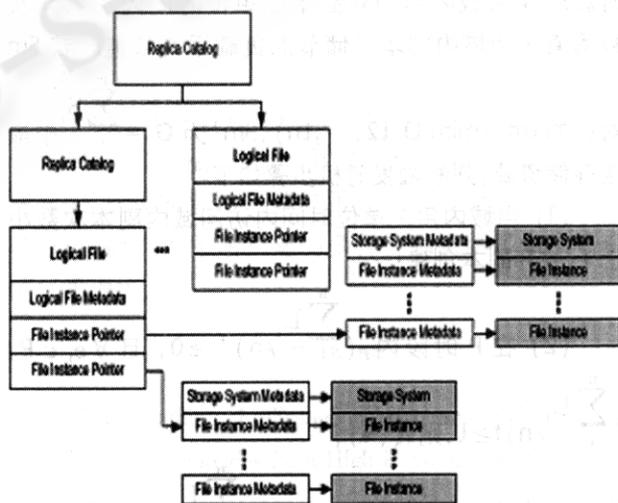


图 1 网格副本结构

### 3 副本创建的策略

#### 3.1 副本创建策略的原则

副本创建策略首先要考虑的是以下三个问题: 创建什么样的副本, 何时创建副本, 在哪创建副本。就是要确定创建副本的时机、地点和内容。策略评估规则必须考虑运行系统负载、存储终端效率、网络状况和数据副本大小等物理特性因素, 并结合用户访问特征决定副本创建时机和地点。

为了更好的控制副本创建, 副本模型根据网络拓扑引起的延迟区分为域内副本衍生和域间副本扩展策略。域内副本衍生策略通过在域内衍生副本增加用户的数据访问点以实现域内文件主机的负载均衡, 域间副本扩展策略依据远程用户访问副本次数使得副本在域间扩展, 减少了用户的访问延迟和带宽消耗。

#### 3.2 域内副本衍生策略

域内副本衍生实现在域内创建多个副本以减轻现有域内数据的访问负荷, 改善负载均衡。域内任务调度器负责将用户对域内数据的访问请求任务分配到域内对应副本宿主机上执行。

设域内有  $G$  个可用存储系统, 其中含同一副本存储系统有  $(S_1, S_2, \dots, S_n)$ , 其对应的负载值为  $(L_1, L_2, \dots, L_n)$ ; 设置参数  $P, Q (P < Q)$ ,  $P$  值要求系统可以对剧烈增长的用户访问请求作出及时响应,  $Q$  值则对副本创建更为谨慎, 其允许文件及副本主机服务器性能暂时波动;  $T$  为域内用户对副本在单位时间内的访问次数阈值;  $L$  为域内副本存储节点负载平均阈值。若  $S_m$

对应的  $L_m = \min(L_1, L_2, \dots, L_n)$ ,  $S_m'$  为  $G - \sum_i^n S_i$  中最佳存储资源, 判断触发时机步骤如下<sup>[3]</sup>:

(1) 当域内用户单位时间内访问域内副本次数小于  $T$ , 放弃副本创建;

(2) 在  $P$  时段内  $(\sum_i^n L_i/n)' \geq 0$ , 且  $\forall \epsilon_t \in P$ ,  $(\sum_i^n L_i/n)t \geq L$  则转(4);

(3) 在  $Q$  时段内, 若  $\epsilon_t \in Q$ ,  $(\sum_i^n L_i/n)t < L$  则放弃副本创建;

(4) 触发副本创建进程, 进程负责启动和监控  $S_m$

到  $S_m'$  的副本传输和创建过程。

#### 3.3 域间副本扩展策略

域间副本扩展实现副本域副本向无副本域扩展的过程。这个过程较域内副本扩展策略更为复杂, 因为在层次网络拓扑结构中, 副本需要创建的位置具有多样性, 根据需求的不同可能会在多个域或者多个层次中被创建。

设网格  $G$  内源文件  $F$  有多个副本  $(R_1, R_2, \dots, R_n)$  分布在网格各个不同的层次域, 其对应的副本数据访问服务范围  $(Q_1, Q_2, \dots, Q_n)$ , 源文件  $F$  的服务范围  $Q_f, Q_f \cup Q_1 \cup Q_2 \cup \dots \cup Q_n = G$ , 且  $Q_f \cap Q_i (i = 1, 2, \dots, n), Q_i \cap Q_j (i, j = 1, 2, \dots, n; i \neq j)$ 。

在上述假设前提下,  $\forall \epsilon_{Q_i} (i = 1, 2, \dots, n)$ , 设  $Q_i$  内包含域  $(D_1, D_2, \dots, D_n)$ , 且  $R_i$  所在的域  $\notin (D_1, D_2, \dots, D_n)$ , 若对应域  $D_i$  内用户访问  $R_i$  的次数为  $h_i (D_1, D_2, \dots, D_{i-1}, D_{i+1}, \dots, D_n)$ , 域用户读访问副本  $R_i$  数据经过域  $D_i$  的总次数为  $h_i'$ , 系统根据副本尺寸大小和  $Q_i$  内网络拓扑结构设定各对应域访问次数阈值  $T_i$ 。

$\forall \epsilon_{D_i} (i = 1, 2, \dots, n)$ , 若  $h_i + h_i' > T_i$ , 则无副本域  $D_i$  向  $R_i$  所在域申请副本创建<sup>[4]</sup>。

### 4 性能对比和分析

由于域内副本衍生的情况比较简单, 因此实验重点介绍域间副本扩展策略的性能对比和数据分析。我们利用 Globus 平台创建的网格环境来比较三种副本扩展策略的性能, 他们分别是: Fastspread、Cascading 和  $T$ -value<sup>[5]</sup>。

假设网格  $G$  中有 7 个主数据域节点, 主节点之间和用户与父结点之间网络带宽都为 100M, 网络延迟为 20ms, 第三层节点域的用户访问位于根节点域的文件  $F_1$ , 文件大小为 100M, 网格语境根据用户的访问特征分为集中式、平均式、分散式三种模式。比较指标有副本存储空间占用、副本访问网络传输时间、副本更新时间三种参数。副本存储空间占用反映文件  $F_1$  的所有副本创建后占用的物理存储空间, 副本访问网络传输时间指副本创建后, 特征用户平均访问一次副本数据在网络间传输的时间, 副本更新采用 Master-Slave 完全机制, 其时间包括源文件的一次更新扩展到所有副本所经历的总时间<sup>[6]</sup>。

### 4.1 集中式

在集中模式下,对文件 F1 的访问集中在某一个用户(如 user1)或某一个用户域(如 D1),在此模式下三种策略副本创建如图 2 所示。

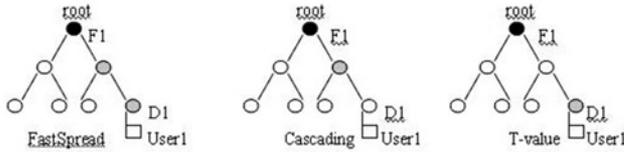


图 2 集中模式副本创建

表 1 集中式副本创建性能指标表

策略	副本存储空间	副本访问网络传输时间	副本更新时间
Fastsread	200M	1.02s	2.04s
Cascading	100M	1.04s	1.02s
T-value	100M	1.02s	1.04s

### 4.2 平均式

在平均模式下,对文件 F1 的访问量平均分布在各底层用户域(如 D1, D2)。在此模式下三种策略副本创建如图 3 所示。

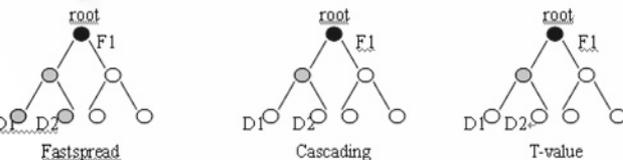


图 3 平均模式副本创建图

表 2 平均式副本创建性能指标表

策略	副本存储空间	副本访问网络传输时间	副本更新时间
Fastsread	300M	1.02s	3.06s
Cascading	100M	1.04s	1.02s
T-value	100M	1.04s	1.02s

### 4.3 分散式

在分散模式下,对文件 F1 的访问量在部分用户域呈现平均趋势(如 D1, D2),在部分用户域呈现集中趋势(如 D3)。在此模式下三种策略副本创建如图 4 所示<sup>[7]</sup>。

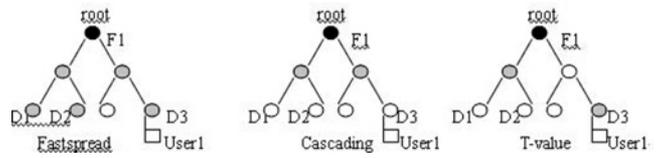


图 4 分散模式副本创建图

表 3 分散式副本创建性能指标比较

策略	副本存储空间	副本访问网络传输时间	副本更新时间
Fastsread	500M	1.02s	5.10s
Cascading	200M	1.04s	2.04s
T-value	200M	1.03s	2.06s

## 5 实验结果分析

从上述特定语境下的对应副本创建性能分析可以看出 Fastsread 策略总是能保持最快的访问响应时间,但其是以牺牲存储空间和更新时间为代价,千兆副本创建存储空间消耗和副本更新时间比其他两种策略的大致 2-3 倍,而千兆副本读访问时间相对于 T-value 却只能减少了 2%。T-value 在保持和 Cascading 的相同存储空间占用和较低的带宽消耗的情况下,副本更新时间虽略有上升,千兆副本读访问时间却有减少[8]。三种策略在需求的侧重面有所不同,用户可以根据具体的访问要求来制定对应的扩展策略。

## 6 结语

网格副本是否能够被正确的创建是网格性能的一个重要方面。不同的访问要求需要设定不同的副本创建策略,因此策略的正确与否很大程度上决定了网格的性能高低,而策略创建的多样性则表明了该项研究的任重道远。

### 参考文献

- 1 K. Ranganathan, I. Foster. Design and Evaluation of Replication Strategies for a High Performance Data Grid. International Conference on Computing in High Energy and Nuclear Physics, Beijing, September 2001.

(下转第 119 页)

(上接第 47 页)

- 2 Byoung - Dai Lee ,Jon B. Weissman : An Adaptive Service Grid Architecture Using Dynamic Replica Management. GRID 2001 63 - 74.
- 3 陈志刚 ,刘安丰 ,熊策等. 一种有效负载均衡的网格 Web 服务体系机构模型. 计算机学报 2005 28( 4 ) : 458 - 466.
- 4 Ming Tang , Bu - Sung Lee , Chai - Kiat Yeo et al. Dynamic replication algorithms for the multi - tier Data Grid. Future Generation Computer Systems 21 2005 : 775 - 790
- 5 D. G. Feitelson , L. Rudolph. Metrics and benchmarking for Parallel job scheduling : Proceedings of the Workshop on Job Scheduling Strategies for Parallel processing , 1998 , 1 - 24
- 6 Kavitha Ranganathan and Iran Foster : Identifying Dynamic Replication Strategies for a High - performance Data Grid. Department of Computer Science , the University of Chicago , Springer - Verlag Berlin Heidelberg , 2001 80 - 85.
- 7 李伟 ,徐志伟 ,卜冠英等. 网格环境下一种有效的资源查找方法. 计算机学报 ,2003 ,26( 11 ) :1546 - 1549.
- 8 David G. Cameron ,Rub en Carvajal - Schiaffino , et al. Evaluating Scheduling and Replica Optimisation Strategies in OptorSim. University of Glasgow , ITC - irst , Via Sommarive 18 , 38050 Povo ( Trento ) , Italy CERN , European Organization for Nuclear Research , 1211 Geneva , Switzerland. 1 - 5