

空间信息网格中的复制技术^{*})

罗滔 方涛

(上海交通大学图像处理与模式识别研究所 上海200030)

摘要 分布性是空间信息网格的重要特点之一。组成网格的高性能、计算能力的不同计算机及各类数据库等各种设备和资源分布在不同的地理位置,跨越的地理范围较广。为了让用户更快更好地获得所需的空间信息,采用动态复制技术,把数据文件复制到用户最容易访问的服务器或数据库。把相同的副本放在地理位置不同的站点,能够减少数据访问的响应时间,增强分布式运算的健壮性。本文将依据 Globus 中的复制体系架构,根据空间信息网格在城市空间信息共享服务与在线分析处理中应用的具体特点,在分析各种复制策略优缺点的基础上,从城市空间信息应用服务的实际特点出发,提出我们的动态复制策略与技术。

关键词 空间信息网格,复制,动态复制策略,复制目录

Replication in Spatial Information Grid

LUO Tao FANG Tao

(Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030)

Abstract Distribution is one of the most important characteristics of spatial information grid. Various settings and resource such as computers, which are high-powered and have different computational abilities, as well as different databases, the electronic libraries and so on, constitute Grid. They distribute in different geographical locations and spread widely. To satisfy the user's need and to make them get what they need fast and reliably, we use dynamic replication technology to replicate files to some server or database that can be accessed easily by users. Identical replicas of data are generated and stored at various globally distributed sites. Replication can reduce data access latency and increase the performance and robustness of distributed applications. According to replication infrastructure, which is used by Globus, and the characteristics of spatial information grid, which is used in the city spatial information service sharing and on-line analysis, we brought forward our dynamic replication policy and technology. They are exalted from the analysis of different replication policies, and found on the practical characters of city spatial information application service.

Keywords Spatial information grid, Replica, Dynamic replication, Replica catalog

1 前言

网格是一个集成的计算和资源环境,或者说是一个计算资源池,它能充分吸纳各种计算资源,并将其转化成一种易获得、可靠的、标准的、经济的计算能力。除了各种类型的计算机,这里的计算资源还包括网络通信能力、数据资料、仪器设备等各种相关资源。

为了满足网格的基本要求,在数据网格中,必然涉及到大量数据文件的复制和传输。复制是通过创建同一个数据的多个拷贝,并通过网络分布到另外一个或者多个地理位置不同的系统中,来提高数据资源的可用性并且可以提高整个网格的有效性和数据资源的使用效率。使用复制策略具有减少存取延迟,减少带宽占用,均衡网格负载,提高数据的可用性等等优点。城市空间信息应用中的数据具有下面的特点:涉及的数据量庞大,往往能达到千兆甚至百万兆字节;不同部门获取的空间信息以分布式方式存储,一般都是存储在通过广域网相连的不同存储设备上;应用对数据资源访问频繁;数据资源的使用者数量庞大,而且分布在不同的地理区域。数据的这四

个特点决定了绝对不能够完全通过网络来远程存取数据文件(否则整个网格性能将是不可忍受的),必须采用复制的手段来将一些数据文件复制到网络中的不同节点上。

既然存在复制,那就存在由谁在什么地点什么时间决定复制什么文件到什么地方上去的问题。对这个问题回答可以分为静态复制策略和动态复制策略两种。如果采取静态复制策略,只是在网络初始状态时静态手工复制,放置一些副本在网络的各个节点上,虽然也能有复制的上述优点,但是它存在静态复制策略不能适应用户行为的变化致命缺陷。最初复制的文件可能稍后根本不再受欢迎,再也没有对它进行存取过。而后来网络上最受欢迎的文件却根本没有复制,致使造成拥有这个文件的站点过量负载,不但浪费了大量的网络带宽,还成为整个网格的瓶颈;同时还要考虑到副本在不同地理位置的站点的分布状态,因为有的数据资源在某个地区需求量大,在某些地区却可能是完全不需要的;而且城市空间信息的内容又是不断更新变化的,还需要把它的变化及时地反映到各个相关站点。但是空间信息数量巨大的数据及文件,手工监控并维护其状态是不现实的,必须采用动态复制策略。动态复

^{*})本文研究得到国家863项目(No. 2002 AA 134020-05)的资助。罗滔 硕士生,主要研究方向为信息网格及其应用研究。方涛 副教授,主要研究方向为网格技术与信息共享、遥感与GIS等。

制策略可以适应用户行为的变化,由动态复制策略来自动决定什么时间复制什么文件到什么地方去。

在国际上,数据网格中动态复制策略已经成为了一个研究的热点,很多国家都开始对其研究。Globus 系统主要针对大型远程数据文件的访问。Globus 复制管理的功能主要包括:创建全部或部分的文件集合的新拷贝,注册新的拷贝到复制目录中以及允许用户和应用程序查询复制目录来找到所有存在的,部分或全部文件的拷贝。复制管理的一个核心部分:复制目录服务(Globus Replica Catalog)通过把部分相关数据放在离应用程序最近的位置,可以快速地访问数据。目前,美国 NASA IPG、欧洲数据网格、美国国家技术网格 NTG、Gr-Phyn、PPDG、ASCII Grid、日本的数据农场 Data Farm 等项目都采用了 Globus 系统。

依据 Globus 中的复制体系架构,根据空间信息网格在城市空间信息共享服务与在线分析处理中应用的具体特点,在分析各种复制策略优缺点的基础上,从城市空间信息应用服务的实际特点出发,提出我们的动态复制策略与技术。

2 各种复制策略

任何动态复制策略都需要回答三个基本的问题:什么时候复本应当被创建?哪一个文件应当复制?复本应当放置到什么地方去?对这三个问题的不同回答产生了不同的复制策略。复制的决定(什么时候应该产生复制)应该依据下面的事实做出:对于当地的计算元来说,复制(包括文件传输和文件删除的花费)是否能够缩减未来文件存取的花费。对于动态复制策略的评价一般是通过平均响应时间和整体带宽占有量来进行的。目前有这样几种常用的动态复制策略。

无条件复制,最老文件删除:只要任务一提交,这种策略总是把文件复制到站点上。假如站点满了,没有空间再来容纳新复制的文件时,在存储单元上面最老的文件将被删除。

无条件复制,使用最少的文件删除:只要任务一提交,这种策略总是把文件复制到站点上。假如站点满了,没有空间再来容纳新复制的文件时,在这段时间内被最少访问的文件将被删除。

最佳客户端策略:网格中的每个节点对于它上面的每个文件都维护着一个详细的请求记录历史表。这个表上面记载了每个请求是从哪个节点来的,它要找哪个文件。最佳客户端复制策略按照下面的方式工作:在一个给定的时间间隔里面,每个节点都检查它上面所有文件的请求记录,看看对于某个文件的请求数量是否超过了它的极限。对于那个超过请求数量极限的文件,最佳客户端就是其中产生请求数量最多的站点。这个节点接着就把那个超过请求数量极限的文件复制到最佳客户端上。一旦它在最佳客户端上创建了一个复本,在服务器端,对于那个文件的请求记录细节将被清空。通过这种方式,凡是超出请求数量极限的文件将被复制到网格的其它地方去。

瀑布复制策略:对这个策略的最好的模拟就是一个三层的瀑布。水从最上层产生,当它注满了最上层的边缘,它开始向下一层流。当这一层也被注满时,那它就流向最下面一层。在这个策略里面,数据以类似的方式流动。一旦对于顶层文件的请求超过了极限,就在下面一层创建复本,注意是在下面一层的最佳客户端创建复本。一旦对于第二层文件的请求超过了极限,那么将把文件复制到下一层面上去,依此类推,也许一个受欢迎的文件将最后被复制到客户端自身那里。

空间信息网格涉及的数据量大,如果采用无条件复制策略,那些没有太多用处的文件和数据也放到各地的存储系统,会占用磁盘空间,造成不必要的浪费和管理上的不便。所以根据用户的请求来确定是否有必要在当地磁盘上复制文件以及被复制的文件的种类数量,并在一段时间以后,将使用最少的文件删除。这样的复制策略,是结合“最佳用户策略”和“无条件复制,使用最少文件删除”两种策略的特点所提出的动态复制策略。

在复制中,复制管理服务是很重要的一部分。它的一个主要功能就是对文件进行操作,包括:(1)注册文件,即把一个复制目录(COLLECTION)上已知的存储系统(LOCATION)上的文件信息记录下来,不拷贝文件本身;(2)发布文件,即把一个复制目录未知的存储系统上的文件信息记录下来,由于复制目录中没有文件的所在存储系统的记录,所以需要把文件拷贝到一个目录已知的存储系统中,并进行记录;(3)拷贝文件,即把对应统一逻辑集合的文件在复制目录已知的存储系统之间进行拷贝;(4)删除文件,即把文件信息从复制目录中清除,也可以选择把文件从对应的存储系统中删除。

使用 URL 来描述远端文件名、文件的物理地址和访问的协议,这样就可以把文件的服务器的位置显示地表现出来,然后应用程序用某些启发式信息对不同数据库服务器上的复本进行选择。返回给用户所选出最佳复本的 URL 地址。用户可以用此 URL 地址,直接连接到想要的文件。

3 复制服务流程

用户不需要知道所需文件的确切的名字或所在的位置。那么,如何获得所需要的文件或数据呢?首先,要详细地描述所需数据的性质,并将属性的描述传输到复制目录表(相当于元数据目录)。复制目录表根据程序提供的属性描述,查询包含这些目录的索引,产生一系列包含描述特性的数据的逻辑文件列表。

接着,复制目录表返回这个逻辑文件列表,把逻辑文件名传给复件(将复制的文件称为复件,下同)管理服务。复件管理服务返回所有注册了的所要逻辑文件的拷贝的物理位置。

第三,把复件的物理位置(URL)列表给复件选择服务,由选择服务确定源和目的存储系统的位置并给所有候选的数据传输操作。

第四,复件选择服务根据各候选者传输性能的评估,基于网格测试和预测,为特别传输选出最好的位置,并返回被选中的复件的位置信息到用户。最后,用户根据返回的 URL 地址可直接连接到它所要的数据或文件。

用户得到自己所需要的数据或文件的详细流程如图1所示。

4 创建管理复件的数据库

从复制服务的流程就可以看出,复制目录在这儿发挥着重大的作用,我们的首要任务就是建立起一个详细完备、可供查询的复制目录表(COLLECTION),并由它来管理当地存储系统管理表(LOCATION)。实际上,为了最大可能地满足用户需求,COLLECTION 还会动态去寻找用户所请求的文件。在每个 COLLECTION 表中都记录了与自身相联系的别的 COLLECTION 表。当它自身的信息不能满足用户需求时,会向别的 COLLECTION 发出请求。这时相对于它所请求的 COLLECTION,它相当于是用户。当请求得到了回应,它会

所得到的文件或数据的 URL 地址传给向它发出请求的原用户。并把此文件相关信息添加入表,再把更新的结果传递到 LOCATION。如果此文件被请求得太多,考虑到带宽和负载

平衡,最佳的用户请求响应时间等因素, COLLECTION 将把文件实体拷贝到由它管理的 LOCATION 中。图2简要地表明了复制目录表和当地存储系统管理表的关系。

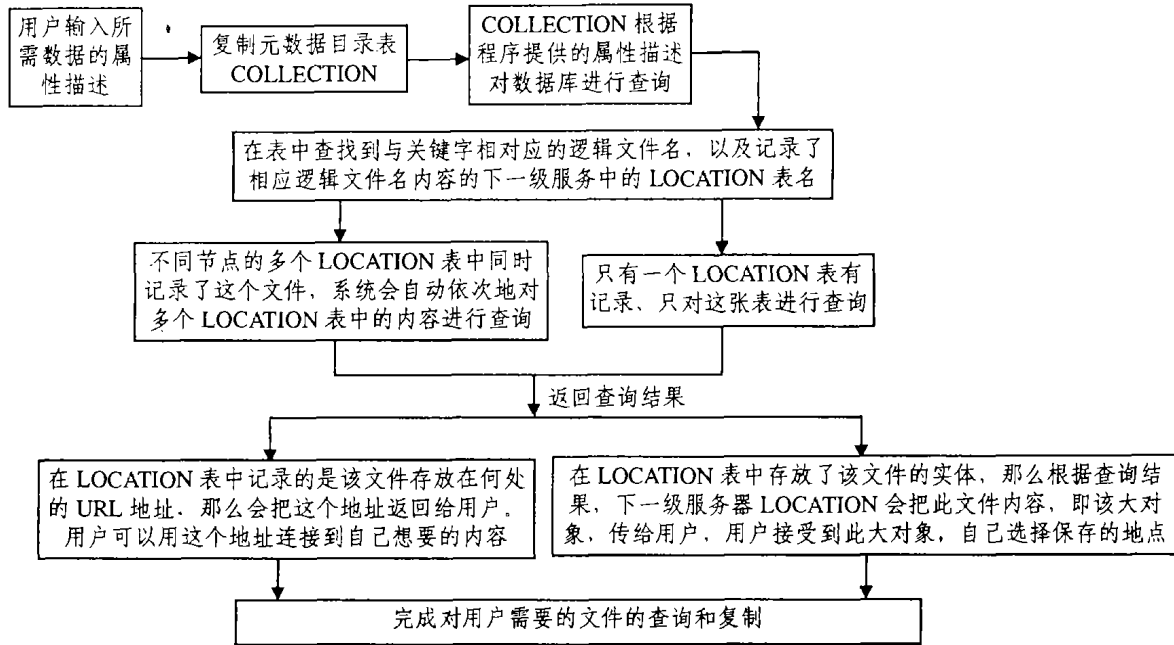


图1 复制流程图

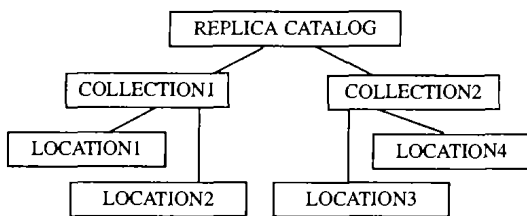


图2 复制的体系架构略图

在当地存储系统管理表中包含了一些复制目录表的内容,这样可方便 SQL 语言直接查询。在当地存储系统管理表中不仅记载逻辑文件名以及它们的属性,文件存放的 URL 地址,提供逻辑文件名到文件实体的映射,也可以保存文件实体。用监测器进行监测,用户对文件(包括逻辑文件名和文件实体)的每一次访问都会刷新“最近访问时间”的记录。当所存放的文件和数据占满了当地存储系统的存储空间时,根据表中“最近访问时间”的记录,最近一段时间使用最少的文件和数据将被删除。

5 复件目录的更新技术

针对上述的需求,我们提出了这样一种解决方案,也就是采用 Oracle 的快照复制。在 Oracle 中,快照是对其他表,尤其是远程表的一个查询拷贝,快照的目的是:只需拷贝一次,就可以多次使用远程表数据。为防止数据变得陈旧无效,可以指定数据刷新的时间,在刷新以前,快照中的数据一直保持有效。当快照第一次刷新之后,即使原表损坏或关闭,只要快照存在,就可以继续使用快照中的数据。

为了能把位于各个地区的当地存储系统 LOCATION 表快照中的变化及时的反映到 COLLECTION,可以使用数据库触发器。数据库触发器是在对表进行操作时最常使用的约束检查工具,当向一个表发出的 SQL 语句满足某个触发条件时,系统便自动激活触发体,执行程序员设定的操作。

结论 本文研究和讨论了在城市空间信息网格中如何运用复制策略与技术,使用户能更快、更好地获得所需的空间数据和分布式的网络资源。所采用的复制体系构架和复制服务增强了对复件的管理。随着国际上先进网格技术及其应用不断取得进展,对如何选择最合适的文件和数据源,进行文件的高效传输,以及复件在分布式网络环境上的最优分布等问题需要进行深入研究,以达到资源利用的最大化和最优化。

参考文献

- 1 Hoschek W, et al. Data Management in an International Data Grid Project
- 2 Allcock B, et al. Data Management and Transfer in High-Performance Computational Grid Environments. <http://www.globus.org/research/papers/dataMgmt.pdf>
- 3 CERN, European Organization for Nuclear Research, Geneva, Switzerland Institute for Computer Science and Business Informatics, University of Vienna, Austria, Distributed Database Management Systems and the Data Grid
- 4 Ripeanu M, Foster I. A Decentralized, Adaptive Replica Location Mechanism
- 5 Stockinger H, et al. File and Object Replication in Data Grid. <http://www.globus.org/research/papers/FileRepCluster02.pdf>