

基于 iRODS 的分布式文件管理共享系统

GDocument

A Distributed File Management and Sharing System Based on IRODS: GDocument

周莹^{1,2} 黎建辉¹ 侯玉芳¹ (1.中国科学院 计算机网络信息中心 北京 100190;

2.中国科学院 研究生院 北京 100190)

摘要: 随着科学数据库中的资源迅速增加,对这些资源的共享使用和管理提出了新的要求。iRODS (i Rule Oriented Data Systems) 是美国圣地亚哥超算中心 (SDSC) 开发的一个数据网格中间件及其相关的工具系统,介绍了 iRODS 的体系结构和主要特点,并基于 iRODS 设计开发了一个分布式文件管理与共享系统 GDocument,它能有效管理分布式异构数据,并提供统一的访问和集成共享。

关键词: 数据网格 分布式 数据集成 SRB Irods GDocument

1 引言

随着信息技术的发展,海量的科学数据已成为现代科学研究和应用领域的重要资源,例如全球气候模拟、高能物理、生物计算、大规模的信息和决策支持系统等的应用,其涉及的数据量将达到几十 T 甚至上 P 的级别,这些海量的科学数据通常来自于不同的研究机构和不同的研究项目,具有分布式、异构式和自治式的特征。现有的数据管理体系、方法和技术已经不能满足人们对高性能、大容量分布存储和分布式处理的需求,因此在计算网格的基础上人们提出了数据网格的概念,以解决信息化科研环境中海量数据的管理、统一访问和集成共享的问题。

数据网格是一种软件系统,它为管理分布的数据记录提供通用的软件结构,并可以支持所有类型的数据管理环境。数据网格的一个重要思想就是在广域分布式环境中对大容量数据进行有效的管理和传输。

数据网格支持数据虚拟化,可以管理分布在多种存储系统上的共享数据集合的属性和数据访问。数据虚拟化是把存储系统的物理属性映射到面向用户数据

的逻辑视图上。共享的集合用来为数据共享、数据发布和数据保存组织分布式数据。

2 数据网格中间件系统iRODS

SRB 和 iRODS 是美国圣地亚哥超算中心(SDSC) Reagan Moore 领导的团队通过近 10 年的开发形成的数据网格中间件及其相关的工具系统^[1]。

2.1 SRB 总体介绍

SRB 主要由服务器和客户端构成,其中服务器包括 SRB server 和 MCAT server,可以运行在 Unix, linux 和 Windows 平台上,客户端有 inQ, MySRB 和 Scomman d。SRB 的核心部分是 MCAT^[2]。SRB 的基本结构如图 1 所示。

2.2 iRODS 的改进之处

基于规则的数据系统 iRODS 是为建设下一代数据管理网络基础设施而研究的项目。它是一种新的分布式数据管理方法,沿用了 SRB 关于数据对象、资源、用户和元数据管理的方法。在 SRB 服务器级别上采用了硬编码的策略来管理数据。

① 基金项目:中国科学院信息化建设专项(INFO-115-C01)

收稿时间:2009-01-15

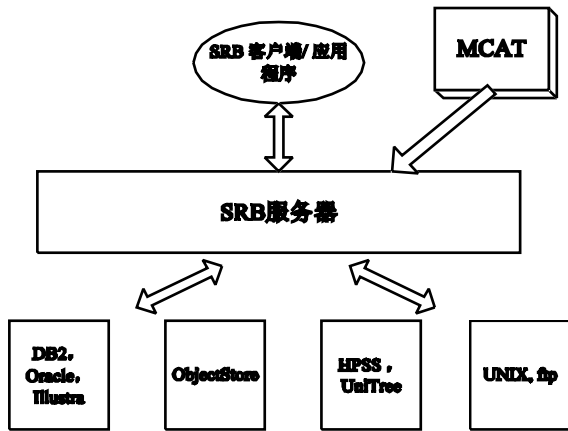


图1 SRB结构示意图

iRODS 是一种可适应的中间件。可适应的中间件体系架构(AMA)能够在不修改程序的基础上满足不同终端用户的需求。我们可以把 AMA 看成透明的玻璃盒子。用户可以清楚地看到系统的工作机制，并启动控制机制来满足各自的需求。而通常情况下，中间件对用户是不透明的。系统开发者具有完全的控制权，各项服务都具有严格的配置选项，用户无法定制，除非系统允许更改中间件的初始配置。在 iRODS 中，开发者用基于规则的编程(ROP)方法来实现可适应的中间件体系架构。基于规则的程序开发不同于普通的程序开发。在 ROP 中，用户更多的控制了功能而非系统和应用程序开发者。因此，用户可以轻而易举地改变、测试和部署特定的过程或者是策略，而不需要系统和应用程序的开发者的任何帮助^[3]。

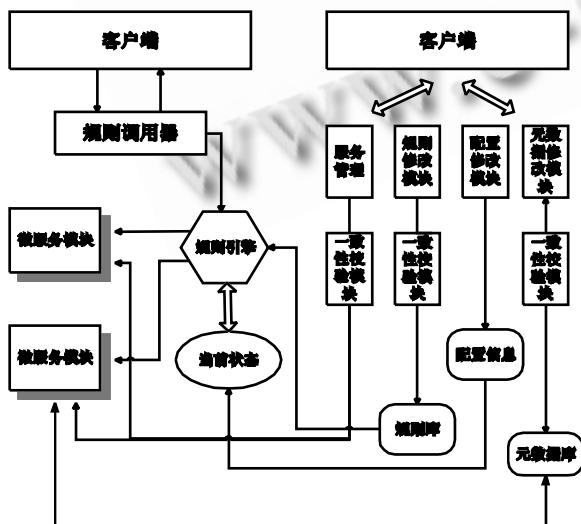


图2 iRODS 体系结构图

如图 2 所示，iRODS 提供客户端和管理端两个访问界面。在管理端，规则库、元数据库、配置文件以及各种类型的微服务模块作为 iRODS 体系的最底层，通过实施一致性校验模块，分别对顶层的服务管理、规则修正模块、配置修正模块以及元数据修改模板提供服务。在客户端，规则引擎作为链接客户端与管理端的一个“桥梁”，是客户端的规则调用器、微服务模块和客户端当前状态描述的连接点。规则调用器一方面要接受来自客户端的资源调度请求，另一方面又需要将通过规则引擎处理后的结果返回给客户端。规则库和规则引擎是 iRODS 的核心组件^[4,5]。

3 分布式文件管理和共享平台GDocument的实现技术

文档是科研最重要的资源之一。各单位积累的学术文档各有侧重，为便于对文档信息浏览、检索，达到资源共享，建立分布式数据管理系统，将文档资源分布到各成员单位，各节点用户可就近访问，上传下载，这样既能减轻中心服务器压力，提高效率，又能提高网络传输的可靠性。本系统是基于 iRODS 内核建立起来的分布式文件型科学数据管理和共享平台，它可以实现广域范围内的异构信息的共享。GDocument 位于客户层和资源层之间，对多种异构存储资源进行统一管理，为客户提供统一的资源访问接口。本系统提供上载用户目录功能，编辑元数据功能，权限管理服务等功能，为实验室不同课题组或者是科学院不同学科的研究者们提供了交流环境，使得文档和数据可以在最大程度上得到共享，并使数据的所有者能够为不同人群提供不同的权限来管理文件。此外，用户还可以基于自己的实际需求制定不同的规则。例如，一个用户希望某个特定集合的所有文件不被删除，包括文件的拥有者和数据网络管理者，然而其他的集合不具备这样的属性。

3.1 GDocument 系统模型架构

本系统的目标用户是实验室的各个课题组和不同学科的研究工作者。本系统可为每个用户建立一个帐户，并用此帐户管理各自的文档。用户可以上传文件目录，编辑文件的元数据，并对其他用户进行权限设置和管理，确定其他的用户对文件的访问级别。

整个系统采用 iRODS 的内核，通过调用其提供的 API Jargon 来开发整个平台。Jargon 包含了本系统的

大部分功能。开发者需要在 Linux 环境下配置好一台带有 iCAT 的 iRODS SERVER 作为整个系统的 Server。iCAT 是类似于 MCAT 的一个由关系型数据库实现的元数据注册系统，主要在进行统一访问时存储和查询系统级和域级的元信息，元信息包括用户、资源、数据聚集、操作许可等信息。可以配置多个不装有 iCAT 的 iRODS Server。这些服务器连在一起可供用户选择上传他们的文档。另外有一个 WEB 服务器，此服务器作为开发平台，当用户想使用分布式文件型数据管理和共享系统时，直接连接登录到 WEB 服务器即可。分布式文件管理和共享系统 GDocument 对终端用户屏蔽了底层信息，用户直接和 web 服务器交互。图 3 显示了整个系统的硬件框架。

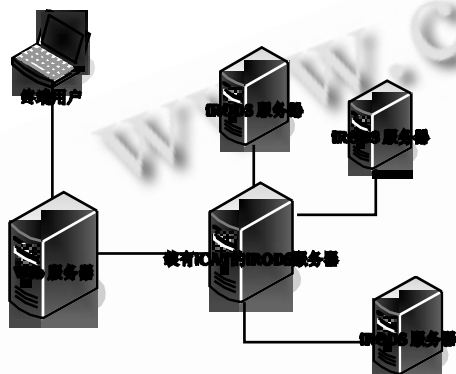


图 3 GDocument 框架示意图

3.2 GDocument 的 web 框架实现

本系统在 Portal 中集成上述文件管理与共享服务，采用 Portlet 与用户交互来把用户的请求与后台的

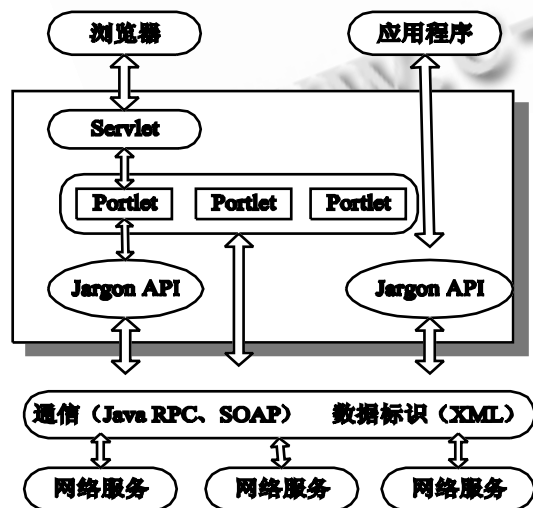


图 4 GDocument 的 portal 结构

web 服务结合，为用户提供了更加便利的手段来访问网络资源。图 4 是 Portal 的结构。

主要的工作有以下几个部分：

(1) 封装服务的调用。在 Portlet 中，对服务的调用都可以封装在一个 Action 里面。比如对分布式文件型科学数据管理和共享系统中的逻辑文件创建 createFile 操作，可以这样做：

```
public class GDocumentAction extends VelocityAction{
    public void createFile(RunData data , Context context){
        GDocumentAction . createFile(filena me , attr);
    }
}
```

(2) 设计与用户的交互界面。下面的代码，显示一个 Form:

```
<form action= " $jslink Template " method=" get" >
    <input na me=" filena me" value=" $fname" >
    <input type=" submit" value=" 提交" >
</form>
```

(3) 注册到 Portal 中，完成部署。注册和部署 Portlet 用的是注册文件，就是用 xml 来描述注册信息。Portal 会读取该注册文件的内容，并且提供出来为用户使用。

3.3 GDocument 的一个实例分析：青海湖基础数据门户服务

为了支持青海湖自然保护区鸟类、生态保护等重要科研，中科院计算机网络信息中心联合青海湖自然保护区、联合研究基地以及相关院所等单位，开发“青海湖国家级自然保护区联合科研基地数据基础平台”，该基础平台的目的是将保护区科研工作中产生的科学数据和资料进行长久保存从而更好地为科研服务，目前该平台中包括鸟类资源调查数据、候鸟监测视频数据、环湖植被样地数据、保护区生物多样性数据等，由于视频数据容量大、文件数众多，根据实际工作中视频数据按文件存储的特点而且这些视频数据未来可能远距离传输的应用需求，我们决定采用 iRODS 来管理视频数据，首先我们用 iRODS 搭建了一个单域多节

点的数据网格试验环境,并在该环境下应用 iRODS 对青海湖的视频数据进行管理和发布。青海湖视频数据的管理属于中国科学院青海湖联合科研基地基础数据库建设工作的一部分,目前的多媒体数据采集自蛋岛、海心山、三块石和救护中心,时间范围从 2004 年 4 月到 2007 年 8 月,一共有 2 万多个文件,数据总量约 140G,其中包括视频文件和图像文件。经过对数据进行分析并同专业人士讨论后,我们定制了青海湖视频数据的元数据模板,并完成了其中精选视频数据的元数据著录。

另外,我们采用 iRODS 提供的 Jargon API 开发了青海湖视频数据的发布平台,Jargon 是 iRODS 提供的基于 JAVA 的 API,面向开发人员使用,用户可使用它来开发满足自身应用需求的 iRODS 客户端,该 API 包提供了 iRODS 常见数据操作的类使用示例,易于学习,开发人员可将重心放在应用逻辑上而无需关注 Jargon 的内部实现细节。目前该发布平台已经基本完成,并取得了较好的效果:

- 1) 用户可以定制规则,管理文档资源更加方便。
- 2) 因为就近上传,大量文档的上传在局域网内完成,与集中式的文档管理相比,数据传输量大大减少。用户可以就近选择服务器进行文档资源的下载,提高了下载的速度和可靠性。

4 结语

总体说来,采用一个稳定成熟的数据网格管理系统作为基础架构,再在其基础上针对具体的数据使用需求开发应用平台能让我们的注意力更好地集中在数据应用逻辑层而无需关心底层的数据访问、传输等问题。本文采用数据网格中间件 iRODS 的内核,通过调用其提供的 API Jargon 来开发整个平台,并应用于青海湖基础数据门户服务,取得了较好的效果。

参考文献

- 1 Moore R. SRB User Manual. http://www.sdsc.edu/srb/index.php/SRB_User_Manual
- 2 Rajasekar A, Wan M, Moore R. MySRB & SRB-Components of a Data Grid. The 11th International Symposium on High Performance Distributed Computing. Edinburgh, Scotland, July 24 - 26, 2002.
- 3 Rajasekar A, Wan M, Moore R, Schroeder W. A Prototype Rule-based Distributed Data Management System. HPDC Workshop on "Next Generation Distributed Data Management". Paris, France. May 2006.
- 4 Moore R, Schroeder W. Introduction to iRODS. https://www.irods.org/index.php/Introduction_to_iRODS
- 5 Moore R, Schroeder W. Rule Engine. https://www.irods.org/index.php/Rule_Engine