

# 应用分布式文件系统优化广西柳工机械股份有限公司 海量文件型数据管理

## Using the Dynamic Technique to Optimize the Management of Massive File Data

邢海韬<sup>1</sup>, 黄孜祺<sup>2</sup>, 肖 健<sup>2</sup>

XING Hai-tao<sup>1</sup>, HUANG Zi-qi<sup>2</sup>, XIAO Jian<sup>2</sup>

(1. 广西柳工机械股份有限公司, 广西柳州 545007; 2. 广西计算中心, 广西南宁 530022)

(1. Cuangxi Liugong Machinery Co., Ltd, Liuzhou, Guangxi, 545007, China; 2. Computing Center of Guangxi, Nanning, Guangxi, 530022, China)

**摘要:**应用微软公司在 Windows2003 服务器版提供的分布式文件系统(DFS)新功能优化柳工机械股份有限公司文件型数据管理。优化重组后的广西柳工机械股份有限公司海量文件型数据,实现了统一的管理模式,所有的员工只需要进入统一的地址,就可以快速地访问到自己所在单位的数据,同时,也可以实现与其他单位的数据交换,不用再担心找不到对方的文件。

**关键词:**海量数据 文件型数据 分布式文件系统

中图分类号:TP315 文献标识码:A 文章编号:1002-7378(2008)04-0351-03

**Abstract:** Application of the Distributed File System (DFS) provided by Microsoft Windows2003 Server version, the new features Liugong Machinery Co., Ltd. file data management is optimized. After the re-optimization of Guangxi Liugong Machinery Co., Ltd. massive data files, a unified modes of management is achieved. All employees by accessing a unified address, can quickly access their data unit. At the same time, they can fulfill data exchange with other units, and don't have to worry about not finding the documents.

**Key words:** massive data, file data, distributed file system

广西柳工机械股份有限公司(以下简称柳工)是我国工程机械行业的骨干企业,主导产品中装载机系列产品的开发水平、制造规模和市场占有率多年来一直居全国领先地位,产销量近年跃居世界前列。作为一个以技术领先为导向的创新型企业,柳工一贯非常重视 IT 技术在企业中的应用,对信息化体系建设也进行了大量投入,历年来的信息化累计投入超过 5 千万元。目前,柳工已经建成一个内连企业所有业务区域,外接互联网、总部和远程多个分部异地相通、固定和移动用户方便协同的立体、庞大、方便、安全的复合网络体系。长期的计算机推广应用,尤其是在技术部门的长期积累,使柳工存储了海量的文件型数据。由于历史的原因,柳工存储的海量文档散布在多台配置不同、性能不同的服务器中,甚至在个

人计算机上也有众多的文档数据。

文件型数据管理从最初级的计算机应用开始,经历了简单的数据管理模式、单服务器的数据管理模式、多服务器的数据管理模式的 3 个阶段。

最简单的文件数据管理模式是在一个对等网内,每个用户通过自己的计算机来共享所需要共享的各种文件和程序,每个用户访问别人计算机上的数据时,必须独立的输入用户名和密码,这样的数据管理模式只适用于不超过 6 人的工作小组,无法适应更大范围的协同工作。

单服务器的数据管理模式是通过一台单独的文件服务器来对网络内的用户提供服务,使用网络来实现文件共享,但是随着网络内用户数量的不断增加以及文件数据量的不断增加,单独的文件服务器的容量负荷较重、网络带宽的负荷增长较快等问题也使应用受到限制。在实际应用中,单独服务器的数据管理模式基本上不能实现网络内的海量文件型数

收稿日期:2008-10-06

作者简介:邢海韬(1975-),男,工程师,主要从事网络管理与网络安全研究。

据管理的要求。

从服务器的性能、磁盘容量以及网络带宽的负荷要求上来说,多服务器数据管理模式基本能够实现海量数据共享的要求,但是,它有一个无法避免的问题,就是所提供的服务凌乱而分散、用户可以存取文档的地方较多,要找到一个不确定的文件往往需要花费较多的时间在不同服务器上进行搜寻。这样的问题是普遍存在的,不仅仅是刚接手的用户会存在,甚至使用较多的老用户也难以方便、快速、准确的找到所需要的文档。为了解决这样的问题,通常可以采用一些厂家的桌面搜索引擎来协助解决搜索的问题。许多计算机管理系统厂家纷纷推出了自己的解决方案,但是这些技术对于海量数据的处理在性能和效率上无法令人满意。为此,我们结合柳工自身的实际情况,选用当前先进的动态管理技术(微软的动态分布式文件系统 DFS)优化柳工海量型文件数据管理,其核心是:以管理为基础,对用户隐掉复杂的技术环节,使用户能够快速方便地访问、管理所需要的数据,而不需要知道这些数据分别处在哪些不同的服务器上,同时提供高级的容错和容灾能力。

### 1 柳工海量文件型数据管理方案的目标和要求

柳工根据自己业务的实际应用情况,提出了对于海量数据管理方案的目标和要求:(1)用户应该能够快速、方便地访问和管理物理上跨网络分布的文件;(2)用户在访问文件的时候要尽可能的简单,不需要记忆过多的位置和步骤;(3)用户可以方便的按照一定的逻辑和要求进行搜索和查找文件;(4)各事业部/部门的文档管理应该能够自行管控,尽可能减少对管理员的要求,同时又能适宜企业的组织和战略变革;(5)普通用户对于其个人的文件夹自行掌控;(6)应尽可能减少投入,降低维护成本,尽可能兼容现有的复杂客户端环境;(7)存放在企业不同区域的服务器上的数据应能够有效利用,且所提供的服务应具备容灾能力。

### 2 柳工海量文件型数据优化

应用微软公司在 Windows2003 服务器版提供的分布式文件系统(DFS)新功能优化柳工文件型数据管理。

#### 2.1 企业情况分析

分解和分析柳工的实际情况总结出相应的应用需求是:(1)在企业内部尽可能的实现统一的管理模

式,以便人员在不同部门和生产厂之间的流动不影响到知识的传递,所有的信息可以快速有效地依据自己原有的经验查找和使用;(2)考虑到不同管理模式的需要,将部门管理与生产厂管理的架构分开,其主要目的是更好的服务现场需要;(3)管理结构细化到具体的生产厂和部门,既适应企业大的组织架构的变化,又能够快速的管理和管理相关文档;(4)在共享数据的同时,要注重良好的安全保障;(5)对个人文档也实现管理与维护,以减少数据丢失的可能;(6)在服务器进行脱机维护的时候,用户端应可以不间断地持续访问而不受到影响;(7)具有较好的数据重组性;(8)可以通过相应的技术手段查找和减少冗余数据的存在,同时,能够对各数据应用组织进行容量和文件类型的限制。

#### 2.2 优化重组的企业文档库结构

图1是企业内文档库结构。优化重组后的企业内文档库结构,实现了企业内部海量文件型数据的统一管理,柳工所有的员工只需要进入统一的地址:\lgnet.com\柳工文档库,就可以快速地访问到自己所在单位的数据,同时,也可以实现与其他单位的数据交换,不用再担心找不到对方的文件。

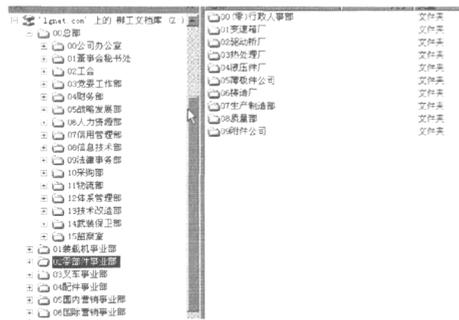


图1 柳工内文档库结构

#### 2.3 优化中可能出现的问题及其解决方法

在使用 DFS 优化柳工海量文件型数据的过程中出现了一些问题,我们进行了相应的解决。

##### 2.3.1 无法使用 DFS

在 FAT 文件系统上安装 DFS,DFS 将无法使用。这是因为 DFS 仅仅能与 NTFS 文件系统一起使用,所以安装时要注意文件系统类型。

##### 2.3.2 管理员无法初始化文件复制

在应用 DFS 优化海量文件型数据的过程中,有时会出现管理员无法初始化文件复制问题。这时需要检查如下几点:(1)通过文件复制服务提供的自动文件复制仅适用于域 DFS,独立 DFS 无法启动 FRS 服务;(2)一个域 DFS 根目录或链接必须至少有两

个目标,才能进行文件复制;(3)不要在由远程存储管理的卷上创建文件复制服务(FRS)副本集,因为会严重影响性能;(4)不要在DFS根目录上启用复制;(5)使用杀毒软件或碎片整理软件会影响文件复制完成。

### 2.3.3 单个服务器负载过大

应用DFS优化海量文件型数据时,也会出现单个服务器负载过大的情况。这是因为维持服务器的负载均衡,需要进行服务器信息同步,所以在完成同步计划时,要综合考虑参与服务器的拓扑、可用带宽、可能产生的复制通信量以及分布式文件系统的复制计划是否合理。

### 2.3.4 用户端数据无法及时更新

应用DFS优化海量文件型数据时,还会出现单个服务器负载过大的情况。这是因为DFS命名空间中可能未包括常用网络共享,这样会导致数据丢失。所以对于那些指向具有时常更改内容目标的链接,应该使用较短的缓存超时值以确保客户端保持有更新的目标列表。

## 3 结束语

柳工在企业的管理运行实践中采用微软的DFS这一先进的管理技术,实现了海量文件型数据管理在多网络、多客户端的复杂环境下的优化,并且所管理的数据超过3TB。这一事实充分说明从企业的实际情况出发,选择并应用合适的技术既能够减

少不必要的管理软件投入,又能够提高企业的管理效率,从而使IT管理能够配合企业的管理变革和战略发展,更好的促进企业的发展。但是由于各企业的实际情况有着很大的不同,评估需求也不相同。主要的评估需求基本有:期望随时添加文件服务器或修改文件位置;访问目标的用户分布在一个或多个站点上;大多数用户都需要访问多个目标;通过重新分布目标可以改善服务器的负载平衡状况;用户需要连续地访问目标;组织中有供内部或外部使用的网站。如果某企业或单位对上述评估需求超过2个以上,就可以使用DFS来优化海量文件数据管理。

### 参考文献:

- [1] Sean Convery. 网络安全体系结构[M]. 北京:人民邮电出版社,2005.
- [2] Rand Morimoto. Windows Server 2003 技术内幕[M]. 刘勇,吴克玲,马玉东,等译. 北京:机械工业出版社,2008.
- [3] Mark minasi, Christa anderson, Michele beveridge. Windows Server 2003 从入门到精通[M]. 马树奇,金燕,译. 北京:电子工业出版社,2004.
- [4] 尚晓航. 网络系统管理:Windows2003 篇[M]. 北京:人民邮电出版社,2006.
- [5] 李劲. Windows 2003 Server 网络管理手册[M]. 北京:中国青年出版社,2004.

(责任编辑:尹 闯 邓大玉)

(上接第350页)

嵌入式电子衡器数据传输系统的设计充分利用了单片机嵌入式系统集成度高、可靠性好、免维护的特点,改善了数据远程传输服务中对恶劣环境的适应能力,降低了生产成本,满足了野外设站、无人值守等应用要求。

本系统可以方便地接入国际互联网,具有成本低、速度快、覆盖面广的特性以及准确性、即时性、完整性及客观性等优点,将极大提高物流企业的工作效率和经济效益,有着广阔的应用前景。

### 参考文献:

- [1] 王峰,时良平. 无线局域网接入点数据监测的实现[J]. 今日电子,2005(6):36-37.
- [2] 陈凯迪,邱飞岳,林文斌. 嵌入式无线局域网的设计与

实现[J]. 中国有线电视,2004(11):29-31.

- [3] 聂家发. 关于802.11协议的研究[D]. 哈尔滨:哈尔滨工程大学,2004:6-7.
- [4] Connect One Ltd. iChip CO2128[EB/OL]. [2008-09-30]. <http://www.connectone.com/products.asp?id=40&pid=55>.
- [5] 沈静. RC4算法及其安全性分析[D]. 广州:广州大学,2007:4-5.
- [6] 胡建人,秦会斌,张智丰. 建筑物内WLAN远程无障碍通讯技术研究[J]. 仪器仪表学报,2006 27(6):2027-2029.
- [7] D Andre Ladson. 利用综合技术来扩展无线局域网的覆盖范围[J]. 中国科技信息,2006(6):305-306,308.

(责任编辑:邓大玉)