国家级气象资料存储检索系统监视分系统的设计和实现*

刘昊钰 马 强 常 飙 张 玺
(国家气象信息中心,北京 100081)

摘 要

监视分系统是国家级气象资料存储检索系统运行维护的主要工具。该分系统针对业务流程和系统设备监视需求,融合多年业务运行维护经验,结合用户需求而开发。运行结果表明:该分系统监视内容全面,软件运行稳定,人机界面友好,吸引了有关用户使用。文章介绍了该系统开发的背景、目的、系统结构和若干关键技术,简述了应用前景。

关键词:国家级;气象资料;存储检索系统;监视

引 言

"国家级气象资料存储检索系统"(the National Meteorological Data Storage System, MDSS)是国家"九五"大中型建设项目"短期气候预测系统工程"的重要建设内容之一。MDSS 是集资料收集和处理、数据存储管理及检索应用等多环节的综合应用系统,其支持的业务范围包括实时和非实时、日常业务与科研工作,应用涵盖范围包括气象系统内部和外部、本地用户与远程用户。该系统的建设和使用,将提高气象信息存储管理和服务的现代化水平,为实现气象信息共享提供基础设施。

"国家级气象资料存储检索系统监视分系统"(MDSS-monitor,下称监视分系统)是 MDSS 的重要组成部分,是管理维护 MDSS 实现其高效运行的工具[1]。运行监视分系统的目标主要有两个,一是实时监视存储检索系统的工作状态,保障系统每周 7×24 h正常工作;二是考查系统性能,为调整系统配置,优化系统性能提供依据。

经过一年半的紧张工作,到 2005 年底,监视分系统项目组基本完成了监视分系统设计和实现工作,部分监视程序已稳定工作了半年有余,收集了超过百万条的系统工作记录。使用监视软件已成为保障 MDSS 正常运行的重要手段。

1 软件需求分析

著名的系统监视软件产品,有 HP 的 OpenView,IBM 的 Tivoli 和 CA 的 Unicenter 等。这些软件经过长期发展和广泛应用,技术成熟而可靠,那么这些软件产品能否全面而有效地满足 MDSS 的监视需求,是否必须另行自主开发监视软件?为回答这一问题,项目组进行了详细的软件需求分析,并调研了典型的监视软件产品。

从功能结构角度分析,MDSS 主要由系统设备和业务流程两个部分组成。系统设备包括服务器、磁盘阵列、磁带库和带库管理软件、分级存储管理软件等基础设施部件;而业务流程是数据收集处理、资料存储管理、资料分发等工作流程。

系统设备监视目的是获取设备工作状态,在其接近工作参数极限,或已经工作异常时实时报警。监视条目根据目标对象的特点设置,如服务器监视对象有 CPU 利用率、内存空间利用率、进程数量、重要文件系统(/,/var,/home,/usr,/tmp)利用率、基本系统进程(syslogd,inetd,pure-ftp,sendmail)运行状态等,在配置了高可用的服务器上要监视的对象还包括高可用进程(群集守候进程、系统日志守候进

* "短期气候预测业务系统工程建设"项目资助。
2005-12-24 收到,2006-07-07 收到再改稿。

程、群集逻辑卷管理器守候进程、群集对象管理器守候进程等)。重要的设备工作状态也需要包括在监视条目中,如网络流量等。对磁盘阵列而言,要求监视磁盘、磁盘控制器等部件损坏情况,并提供空间利用率、数据吞吐量等工作状态。

执行业务流程是 MDSS 运行的目的,也是监视的重点。业务流程监视主要是检查工作流程是否顺利进行。根据 MDSS 的工作流程,应用监视项目组成了"节目单",包括实时数据库所有要素资料的定时入库信息、表数据的备份和清除操作等,综合数据库和卫星资料接收等业务的操作也包括在内。

项目组在 MDSS 上部署了 HP 公司出产的 OpenView A.07.23 软件,并分析了其特点。OpenView 是一个用于管理网络、主机、应用及 IT 运行维护流程的工具集。该工具集包含了几百个模块,用来对 IT 部门的各个系统及其运行环境进行管理。项目组选用的 OpenView Network Node Manager (NNM)模块是系统和网络管理的基础平台,具有自动发现和监控网络节点、监测网络、网络故障诊断等功能。Vantage Point 模块是集中的系统管理工具,能自行发现系统中出现的问题,提醒管理员注意。Vantage Point 的中央控制台收集各种事件,如系统管理、网络管理事件以及数据库和中间件管理事件,并且可以定义一些自动执行的动作,当出现问题时能够进行自动处理。Vantage Point 的事件关联服务(Event Correlation Services,ECS)删除不必要的和无意义的信息,标识问题起因,只将影响关键业务的信息提交给管理人员进行分析。经过使用,项目组认为 OpenView 软件有许多优点可以借鉴,但是将 OpenView 投入 MDSS 监视业务需要再次开发;英文版 OpenView 软件用于业务值班操作过于复杂。

经过细致调研,认识到大型平台监管软件的发展方向,不再是简单地针对系统设备的管理,其范围已扩展到应用管理,特别是对业务流程的监管。在电信、金融、制造等行业中已经广泛应用软件进行业务流程的监视,但在气象资料存储检索应用中还未看到相应的较为成熟的软件产品。

同时,监视分系统不仅在国家级气象信息中心有应用需求,它的设计、开发和应用对区域级、省级气象信息中心也具有一定的示范作用;此外,拥有自有知识产权的软件更利于监视分系统的推广。

综合考虑种种因素,项目组决定自主开发监视分系统,设计并实现有效监视 MDSS 系统设备和业务流程,符合气象信息业务运行规范,结合多年业务运行维护经验,能对 MDSS 运行进行评估和报告的软件。

2 软件结构设计

MDSS 是随业务要求而升级的系统,随时有新设备和业务加入。与之对应,监视分系统也需要开放、规模可扩展的体系结构。体系结构设计要确定软件的基本组织和全局控制结构,特别是模块间通信协议、数据存储的定义[2]。

监视软件的功能可以分解为:被监视对象工作数据的获取、数据的分析和存储、数据的表现。因此,软件组织结构采用分布式多层结构(图 1)。

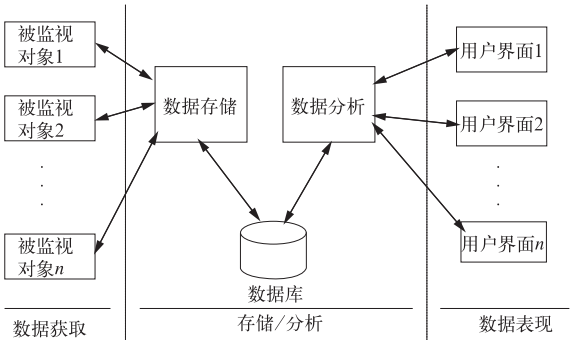

图 1 监视软件的组织结构

Fig. 1 Software Architecture of the monitor subsystem

数据获取层工作在被监视对象上,获取被监视对象的工作状态数据。每类监视对象都有一组监视条目,如服务器类监视对象的监视条目有 CPU 利用率、内存利用率、基本文件系统利用率等。

存储/分析层,运行在"管理监视"服务器上,负责对监视数据的分析,并存储在数据库中。这一层需要分析数据的有效性(监视功能运行后,数据才是有效的),时效性(是否为过时数据),判断监视对象的工作状态是否达到报警门限及报警级别。分析层的另一重要功能是统计系统和业务运行情况,作为系统升级、优化时的依据。

数据表现层运行在监视终端上,为管理人员提供系统运行状态信息,是监视分系统主要的人机界面。

在 MDSS 这样的分布式系统中,层与层之间的数据交换是软件体系结构的关键环节。

数据获取层和存储/分析层之间的接口联接被监视对象和监视软件核心,要求能传递来自硬件(如服务器、磁带阵列)和软件(如数据库、Web 服务器)的监视数据。这一接口,还应该能灵活地纳入或取

消被监视对象。利用关系数据库结构来格式化接口,既实现了与设备无关性,又宜于修改。经过细致讨论,项目组制定了《监视信息库表结构设计规范》,规范中定义的库结构如图 2 所示。

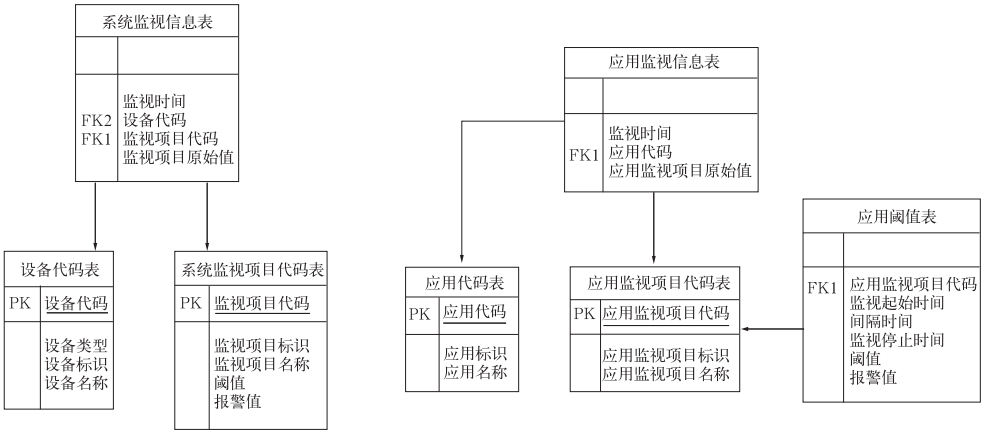

图 2　监视信息库表结构(PK:主键;FK:外键)

Fig. 2　Table structures of the monitor information database (PK: Primary Key, FK: Foreign Key)

存储/分析层和数据展示层之间的接口是将分析后得到的数据向外发布的接口。关心监视数据的用户,不仅包括实时业务运行值班员、维护系统的管理员和关心系统运行状态的业务管理人员,也有与 MDSS 相关的其他监视软件,如国家气象信息中心将要建设的包括各台室业务的整体监视系统。由于不同的用户所关心的数据是不一样的,有的只关心总体运行状态,有的需要具体数据,因此分项目组根据需求分析的结果,设计了树形的访问接口,如图 3 所示。树形访问接口提供了在各层次上了解 MDSS 运行状态的方法。关心整体状态的用户可以访问较高层次的接口,如 MDSS 状态接口,设备状态接口和业务状态接口;而需要具体数据的用户,可以通过较底层的接口取得数据,如实时库服务器状态接口中的 CPU 利用率,实时库的产品资料入库统计信息接口中特定目录下来处理的文件数量。

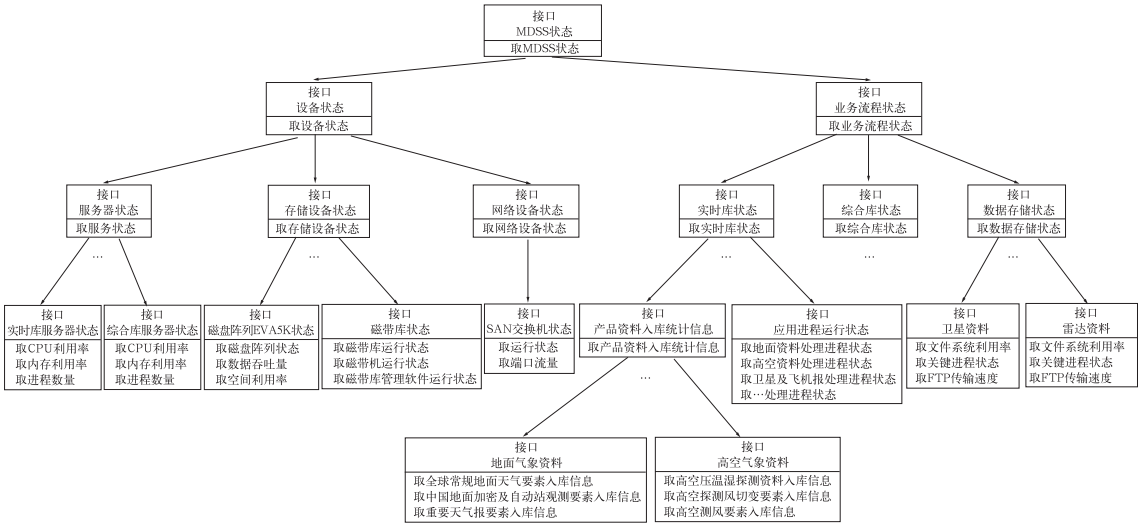

图 3　存储/分析层和数据展示层接口

Fig. 3　The Interfaces between layers of storage/analysis and data view

3　关键实现技术

被监视对象分为系统设备和业务流程。获取系统设备工作状态数据的手段有多种选择。分项目组参考有关商业软件的实现,结合 MDSS 设备的特点采用合理的方法。这些方法有:利用系统命令的输出,如用 HP-UX 的系统命令 bdf 取得文件系统利

用率;利用设备厂商提供的应用程序结果,如采用Legato软件的tapestats程序取得迁移软件工作状态,以及读取日志等。

MDSS业务流程是气象行业独有的,缺少现成的监视手段,项目组必须自己设计。在MDSS业务中,数据库操作是主要的流程内容①。监视分系统必须实现高效、可靠的数据库监视。数据库监视的主要对象有观测资料入库统计信息、国外格点资料入库统计信息、表备份信息、表清除信息等。涉及到表操作、关心表中内容的监视,不宜采用低效率的外部程序定时轮询方式,而采用单纯内置触发器(trigger)又难以实现必要的数据分析功能。项目组采用将JAVA 存贮过程挂在特定数据库表上,组成JAVA 触发器的方法。当表进行插入、删除、修改操作时引发JAVA 触发器动作,JAVA 触发器将此次操作结果分析后反映为业务流程进度。JAVA 触发器的具体实现方法请参阅文献[3]。分项目组实现的JAVA 触发器已经稳定运行了近1年,实践证明Oracle内置的JAVA 虚拟机是可靠的,而且这种方法不影响所依附表的正常操作。

数据展示层是显示监视数据结果的人机界面。为了方便用户使用,分项目组选择通用网络浏览器作为数据展示层的主要容器。传递数据集是通过访问图3所示的树形接口实现的。项目组没有采用现有中间件服务器软件,而是自行开发了中间件服务器。中间件服务器从监视软件内置的数据库中得到系统运行状态数据,经过分析后的数据集驻留在内存中。人机界面需要数据集时,直接访问内存,避免费时的数据库访问。经过比较,这种方式比传统直接访问数据库速度快很多。

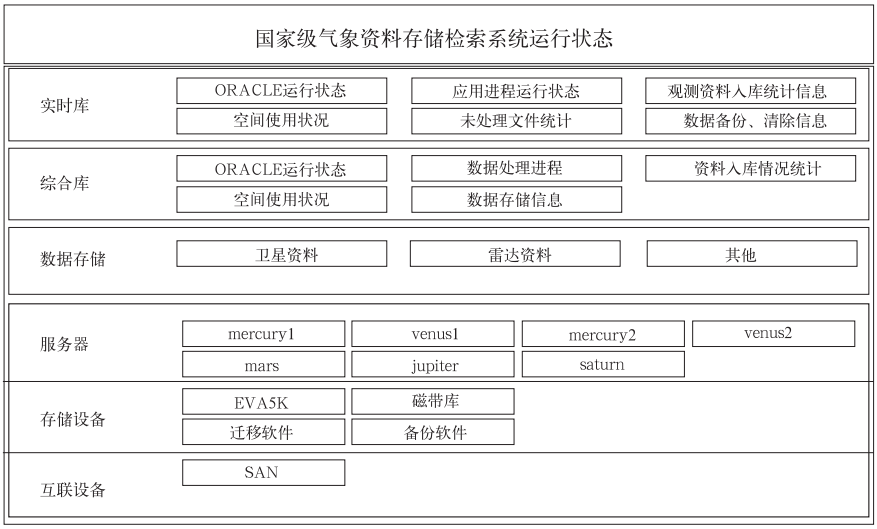

图4 MDSS监视软件人机界面

Fig. 4 The user interfaces of monitor subsystem in MDSS

此分系统的开发涉及了国家气象信息中心多个业务流程,以及各种软硬件设备,工作量大。分项目组在建设和维护国家气象信息中心存储系统的同时,创造性地开展工作,不仅在软件设计和实现技术中取得了经验,在软件开发的组织管理上也进行了尝试。分项目组在软件开发过程中进行了成本控制。控制成本的重要成果是实现了组件的流水线生产方式,这种方式同小规模的程序开发相比明显地提高了生产效率,保证了产品质量[4]。

4 结 语

分系统数据库中保存的数百万条记录,蕴藏了丰富的系统运行知识。利用数据挖掘手段可能取得大量有用信息。

监视信息是发现故障和性能优化工作的依据。例如,通过查看监视信息,项目组发现了系统建设中对SAN交换机的设置失误。系统中的两台SAN交

① 国家气象信息中心. 实时库数据表结构. 2005.

换机设计成负载均衡形式,工作时它们的吞吐量应该大致相同。但是,监视分系统提供的数据显示两台交换机实际的吞吐量差别很大。交换机 A 的吞吐量接近它的工作峰值 200 MB/s,而交换机 B 的吞吐量却几乎为 0(图 5a)。经过检查,发现这是由于某次设备调整时,工程师将磁盘阵列指向交换机 B 的数据路径关闭引起的。打开 B 机的数据路径后,监视分系统提供的数据表明 A 机的吞吐量下降到 120 MB/s 附近,而 B 机的吞吐量上升到 80 MB/s 左右(5b)。

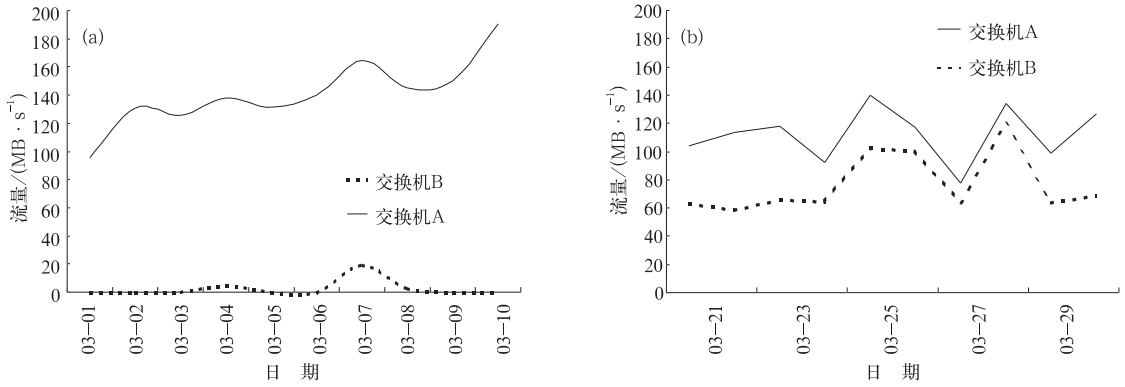

图 5　数据路径调整前(a)、调整后(b)交换机吞吐磁盘阵列数据量比较

Fig. 5　The comparison of switches throughput of disk array data before (a) and after (b) adjustment

MDSS 监视分系统目前已基本实现主要功能,并在保障 MDSS 正常运行,捕捉系统运行"瓶颈"等方面发挥了作用。实践表明监视分系统的设计是合理,实现是可靠的。

参 考 文 献

[1] 沈文海,赵芳,高华云, 等. 国家级气象资料存储检索系统的建立. 应用气象学报, 2004,15(6):727-736.

[2] Mary Shaw, David Garlan. Software Architecture: Perspectives on an Emerging Discipline. New Jersey: Prentice Hall, Inc, 1996.

[3] ORACLE Corporation. Java Stored Procedures Developer's Guide. 2001: 2-17.

[4] Gerald M Weinberg. The Psychology of Computer Programming, Silver Anniversary Edition. New York : Dorset House Publishing Co, Inc, 1998.

The Design and Implementation of the National Meteorological Data Storage System's Monitor Subsystem

Liu Haoyu　Ma Qiang　Chang Biao　Zhang Xi

(*National Meteorological Information Center*, *Beijing* 100081)

Abstract

The purpose of building the monitor subsystem is to guarantee the safe operation of the National Meteorological Data Storage System (MDSS) and to guide the updating of the DMSS. As an infrastructure for sharing scientific meteorological data, real-time operation and post-standardization are supported by MDSS. According to their functions, all components of MDSS are divided into two groups. The supporting equipment includes hardware as well as their accessory software, and the operation of real-time database integrates the database and the data storage. In the equipment group, the targets vital for the smooth running of equipments are monitored, such as the CPU usage and the memory usage of services. Most of

the objects monitored in the group of operation are time-series and sequence operation. For instance, the amounts of observed data are entered into the real-time database at a given time. The possibility of using the current monitor software is also considered, and their shortcomings are presented. The structure of the subsystem is divided into three modules: the collection module, the storage and the analysis module and display module. The monitored operation data are gathered repeatedly from the MDSS by the collection module. The gathered information is calculated by the storage and analysis module to confirm the state of the MDSS and the results are stored in the database and the service memory. The state of the MDSS is shown by the display module on the Web pages. A database scheme is designed as the communication channel between the collection module and the storage module. A tree shape program interface is designed for passing information between the modules of storage and display for users on different levels. Some special techniques which are keys for the implementation of the monitor subsystem are explained, such as the Java Stored Procedure. Experience and the following work of the subsystem are also discussed. The reliability and stability are demonstrated by the performance of the subsystem.

Key words: system requirements; function designs; technique schemes monitor

WMO热带气象工作组会议在广州召开

2007年3月22—24日,世界气象组织热带气象工作组(WGTM)会议在广州召开。来自中国、美国、法国、日本、澳大利亚、马来西亚、印度、中国香港以及世界气象组织的相关官员和专家参加了会议。此次会议由世界气象组织和中国气象局组织、中国气象科学研究院和广东省气象局承办。

热带气象研究工作组是隶属于世界气象组织大气科学委员会的一个专门工作组,该工作组主要职责是通过开展学术交流及推动重大国际计划,促进热带气旋的研究与预报工作,在全球防灾减灾工作中发挥着积极的作用。多年来,中国科学家一直积极参与该工作组的工作,并在其中发挥主导作用。

会议由热带气象研究工作组主席、中国气象科学研究院陈联寿院士主持,广东省气象局局长余勇到会并致欢迎辞。世界气象组织大气研究及环境计划司(AREP)司长 Leonard Barrie 参加了会议,并对中方组织此次会议表示感谢,随后,他详细介绍了世界天气研究计划(WWRP)和 THORPEX 等计划。会议回顾了2006年在南非开普敦召开的世界气象组织大气科学委员会第十四次届会的相关决议。陈联寿院士就WMO大气科学委员会热带气象工作组在2006年的活动做了工作报告。中国气象科学研究院陈德辉研究员对T-PARC计划进行了介绍。会议上,热带气象工作组相关计划负责人也就各自负责的领域作了相关工作报告,其中包括热带气旋研究、季风研究、热带及气候变化对台风的影响数值模拟等。来自世界3个季风研究中心(北京、吉隆坡、新德里)的专家就季风研究活动作了相关报告,报告内容均被纳入热带气象工作组未来发展报告中。会议还就有限区域模拟、热带气象工作组组织机构调整、发展战略及发展目标等问题进行了研讨,并确定了2007—2009年工作计划。

(中国气象科学研究院办公室)