气象计算网格平台资源监视模块的设计与实现*

王 彬 常 飚 朱 江 刘春花
(国家气象信息中心,北京 100081)

摘 要

气象计算网格聚合的计算资源具有地理分布、系统异构、运行状况与使用负载各不相同等特点。气象计算网格平台软件系统的资源监视模块,涉及了远程气象网格节点、资源状态信息获取、web 展示等 3 个层次。资源状态信息获取层可分为轮询、收集、整理等功能,web 展示通过资源地理视图和资源列表栏目实现。基于 ARCON 客户工具箱技术开发实现了资源信息轮询和收集功能。现已实现对国家级网格节点及北京、成都、广州、沈阳 4 个区域级中心网格节点和安徽省级网格节点的 10 个高性能计算机系统的集中监视。

关键词:资源监视;气象计算网格;网格平台软件系统;资源地理视图;计算资源

引 言

从 2005 年底至今,国家气象信息中心联合多家单位及研究机构,建立了一个气象部门内全国范围的、分布的、跨广域网络的气象计算网格,并进行了研究开发,形成了一套网格平台软件系统。

气象计算网格的整体架构由位于国家级主节点、分布在全国不同地区的 8 个区域气象中心分节点、省级节点组成,通过全国气象宽带网络连接起来[1]。现已聚合了分布在国家级、区域、省的 10 个异构高性能计算机系统,总计算能力超过 26TFLOPS,占气象部门内 50%以上。

气象网格平台软件系统可分为计算资源、平台软件、资源管理、应用服务、用户接口等几个层次,包括资源监视、资源动态调度[2]、资源作业查询、用户安全[3]、交互访问、数据支持服务、数值预报模式插件等模块。

气象计算网格整合的资源位于不同地点,系统架构、运行情况、使用效率各不相同。为使网络计算用户、运行值班和系统管理人员了解这些资源的即时状态信息,软件研发中设置了资源监视模块,收集加入气象计算网格的各地气象高性能计算资源的运行状态信息,在 web 门户系统进行集中展示。

1 模块设计

1.1 架构设计

从架构上来看,资源监视模块涉及到 3 个层次:远程气象计算网格节点、资源状态信息获取和web展示。

远程气象计算网格节点:国家级、区域级、省级气象信息中心通过部署、配置、运行了计算网格平台软件后,将其管理高性能计算机系统整合加入到全国气象计算网格,成为一个网格节点。

资源状态信息获取:资源监视模块的主要功能,实现了资源状态信息轮询、收集、整理功能,是网格平台软件系统的组成部分之一。

web 展示:通过 web 页面展示计算资源状态信息,位于web门户网站内,是其动态页面内容的一部分。

1.2 资源状态信息设计

计算资源状态描述了一个气象网格节点内高性能计算机系统信息,可分为 3 个主要部分:概要信息、节点信息、作业信息。

概要信息描述了计算资源的总体状态信息,包

* 公益性行业(气象)科研专项"基于网格的气象计算资源管理技术研究"(GYHY200806018)和科技部基础条件平台国家气象网络计算应用系统建设项目(2005DKA64005)共同资助。
2009-03-17 收到,2009-08-07 收到再改稿。

括资源名称、所属单位、位置、描述、状态、负载等内容。节点信息描述了计算资源内部节点的状态信息。作业信息描述了计算资源上运行的各个作业信息。

1.3　资源状态信息获取流程设计

资源状态信息获取层包括资源状态信息查询、资源状态信息收集和资源状态信息整理 3 个软件子模块和若干信息文件。气象计算网格节点内的高性能计算机系统通常通过一个专用的网关软件服务接入网格。

资源监视模块内资源状态信息获取流程设计如图 1 所示。

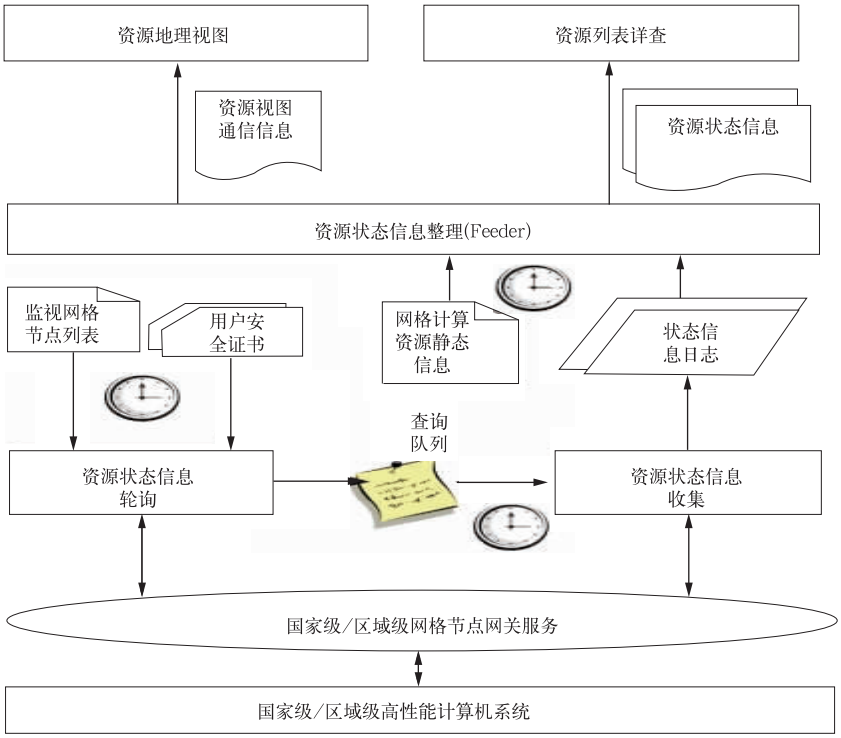

图 1　资源状态信息流程设计

Fig. 1　Design of data flow of resource state

• 资源状态信息查询　按照预定时间和间隔,信息轮询子模块读入预配置好的监视网格节点列表文件,获取相应的安全访问证书,然后通过网格访问协议连接气象网格节点的网关服务,以异步方式提交资源状态信息轮询请求。随后将查询请求信息插入到查询队列文件里。

• 资源状态信息收集　信息收集子模块每隔一段时间定时启动,读入查询队列文件,然后根据读入结果,通过网格访问协议连接气象网格节点的网关服务,获取资源状态信息轮询的结果。无论成功与否,信息收集子模块都会将查询结果写入到状态信息日志文件里。随后更新查询队列文件,清除相应的队列项。

• 资源状态信息整理　信息整理子模块定时启动,读入状态信息日志文件和网格计算资源静态信息,然后转换生成了资源地理视图所需状态消息文件(XML 格式),并主动推送提供给资源地理视图显示,随后转换生成了资源列表详查栏目所需状态信息文件(XML 格式),并主动推送提供给资源列表详查栏目显示。

1.4　资源状态信息显示流程设计

web 展示层直接面向用户,基于通用浏览器的方式展示计算资源的状态信息。web 展示层由资源地理视图和资源列表详查栏目两个子模块具体实现,以不同方式展示计算资源的状态信息。资源地理视图结合资源所属网格节点所在不同的地理位置,在地图上展示资源的概要状态信息,用户可直观、全局地获知气象计算网格内资源分布情况和即时概要状态信息。资源列表详查栏目以列表和图表页面形式给出每一个计算资源全面详尽的状态信息。

2 模块实现

2.1 气象网格节点的构建

网格平台软件系统采用了 UNICORE[4] 技术作为基础平台软件。UNICORE 系统基于 C/S 架构。目前通用的客户端是基于 Java 应用程序开发的可视化用户界面，服务器端由网关服务、网络作业管理服务、目标系统接口服务、用户信息库、具体化信息服务等模块构成[5]。

在国家、区域或省级气象计算(信息)中心部署 UNICORE 软件后，搭建成为气象计算网格节点。UNICORE 网关服务通常运行于一台专用的网格接入服务器上，提供了对网格节点的单一访问入口。一般而言，一个(国家级、区域级或省级)气象信息(计算)中心提供了一个唯一的网关服务，来访问使用这个中心所有的计算资源。

2.2 资源信息轮询和收集功能的实现

监视模块需要连接计算网格节点，进行安全互校验，提交查询脚本，获取资源状态。显然图形化的、"固定动作"的 UNICORE 客户端不能满足程序开发定制运行的需要。经过调研，决定采用 ARCON 客户工具箱技术实现监视模块中的节点访问功能。

ARCON 客户工具箱(ARCON Client Toolkit)[6] 是由 Fujitsu 欧洲研究院支持的一个开源软件项目。这个工具箱提供了访问 UNICORE 网格站点的一个轻量级、简明的 API 编程接口。基于 ARCON 客户工具箱(4.6.1 版本)，开发实现了资源状态信息轮询和资源状态信息收集的代码。

信息轮询通过作业形式提交资源状态查询请求，提交作业可以采用同步或异步方式进行。考虑到各网格节点通过气象广域网络连接，分散于全国各地，属异地分治异构。为了提高监视模块代码的健壮性和适应性，并减少网络资源的堵塞占用，状态信息轮询的代码主要采用异步方式实现。

信息收集通过作业返回结果的方式获取资源状态信息。网格节点返回的结果信息包含两部分：结果对象以及作业特有信息。由于各地区的高性能计算机使用的作业管理软件不同，因此查询结果文件的格式和内容有所不同。

2.3 资源状态信息获取层主要程序包

资源状态信息获取层主要程序包包括基础驱动、作业调度、查询队列日志、资源状态解析、配置文件等 Java 类，实现功能描述见表 1。

表 1 资源状态信息获取层主要程序包

Table 1 Major packages of resource state information acquisition layer

程序包名称	Java 类	描述
基础驱动	nmic.unicoremon	程序的用户接口，启动程序
作业调度	nmic.unicoremon.job	负责提交查询作业和取回结果
查询队列日志	nmic.unicoremon.log	为信息轮询和信息收集的作业日志提供相关接口
资源状态解析	nmic.unicoremon.res	对查询结果进行解析并生成 XML 文件，支持多种作业管理系统的查询结果
配置文件	nmic.unicoremon.conf	负责读取监视网格节点列表、作业执行等配置文件

程序包运行时作为一个系统守护进程常驻内存，启动两个线程，一次查询从一个信息轮询线程发起，提交一个查询作业，记录在查询队列日志中。另一个信息收集线程监听查询队列日志，将查询完成的作业取回结果。资源解析模块启动相关解析，将结果生成统一格式的 XML 文件。两个线程的运行频率可以用配置文件设置。

2.4 资源状态展示功能的实现

资源列表详查栏目采用通用的 J2EE web 组件技术实现，服务器端使用 Servlet 技术获取各计算资源的最新状态信息，并且响应前端的显示更新请求。资源地理视图采用了 Flex+J2EE web 组件技术实现。

Flex 技术是实现具有复杂交互 web 应用程序的一种代表性技术，具有用户友好性、交互性、跨平台兼容性、一次加载多次使用、客户端数据缓存、高效的网络数据信息传输等特点[7]。开发平台选用了 Adobe 公司的 Flex Builder2.0 工具，自主开发了 FlexMap 包，实现了地图上气象计算资源的即时状态显示和交互功能，以直观形式表现高性能计算机资源的物理分布情况，以文字形式展现高性能计算机节点的状态信息。

3　模块部署与运行

随着气象网格平台软件系统在全国各网格节点的部署运行,现已实现了对国家级网格节点及广州、北京、成都、沈阳 4 个区域级中心网格节点和安徽省级网格节点内 10 个高性能计算机系统的集中监视和展示,如图 2 所示。

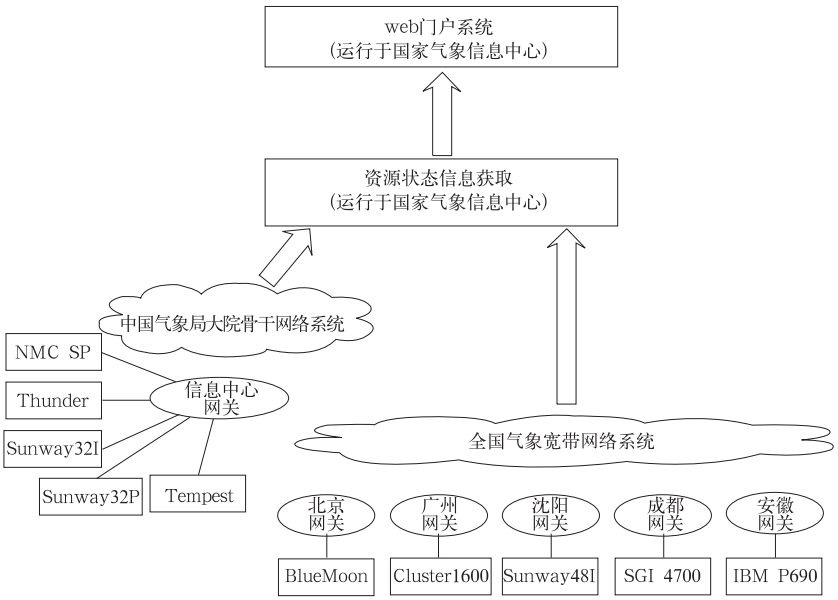

图 2　资源监视模块的部署情况

Fig. 2　Deployment of resource monitor module

资源状态信息获取层的各个子模块运行于国家气象信息中心内。

资源地理视图嵌入到 web 门户系统主页运行,如图 3 所示(绿色表示状态正常)。鼠标停到资源上面时,就会给出资源的即时概要状态信息。

图 3　资源地理视图

Fig. 3　Resource GIS view

资源列表详查栏目如图 4 所示，列出所有资源的状态信息，包括系统名称、所在位置、系统架构、处理器数量、理论峰值、内存、磁盘、负载、作业数及状态等信息。

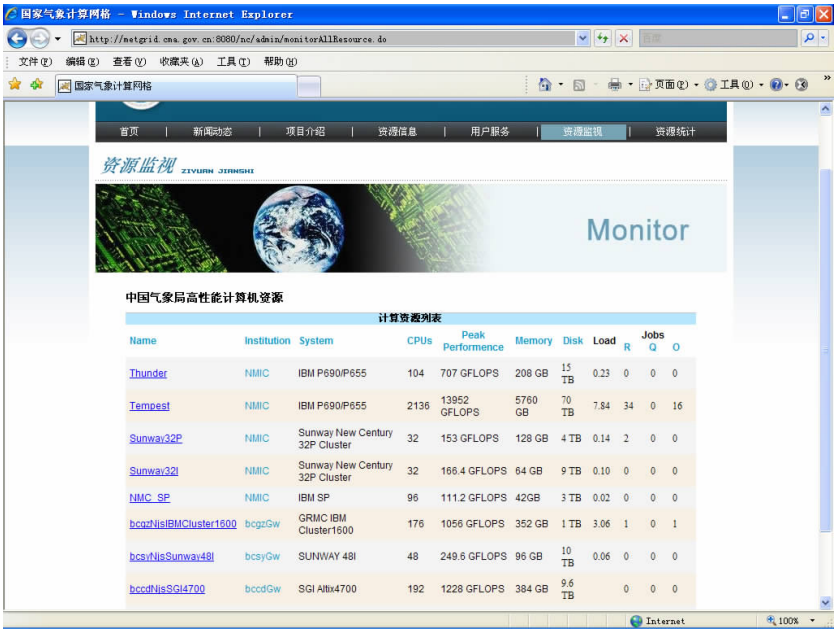

图 4 资源列表详查栏目

Fig. 4 Resource list column

还可以查看某一个资源的详细信息，例如广州区域气象信息中心高性能计算机系统的状态信息如图 5 所示。可以获知系统的各个节点状态、负载、运行具体作业等信息。

此外，应安徽省气象局要求，在国家气象信息中心开发部署了安徽省数值预报模式业务备份系统，一旦发现安徽省高性能计算机系统状态出现问题，立即启动国家级备份系统。为实现安徽省气象局计

图 5 查看单个计算资源的详细状态信息

Fig. 5 Detailed state information of a computing resource

算机系统的 24 h 实时监控,资源监视模块将安徽省气象局 IBM P690 系统的状态信息推送到计算机室监视值班系统内,实现了业务备份监视触发机制。

4　结　语

资源监视模块是气象网格平台软件系统的核心组成模块之一,现已投入业务运行,提供服务使用。资源监视模块能够及时、准确地反映气象网格整合计算资源的运行状态,为气象网格计算用户资源使用和值班工作提供有效参考。同时,监视模块的运行状态也能反映网格中各节点的网络连通和运行状态,为保障网格环境正常运行提供支持[8-12]。

下一步工作将在资源监视模块运行一段时间后,不影响气象网格正常运行的情况下,对监视模块运行状态进行分析,对其性能进行优化,通过调整监视数据采集流程并优化配置等工作,使监视模块占用更少的网络和计算资源完成监视任务[13-16]。

随着气象计算网格的深入建设,资源监视模块将推广到更多的计算资源上,实现对气象部门内所有计算资源的监视。

致　谢:感谢广州、北京、成都、沈阳、安徽等区域/省级气象信息中心对本研究的大力支持!

参 考 文 献

[1] 王彬,宗翔,田浩. 国家气象网络应用计算系统的设计//国家气象信息中心 2007 年度科技年会论文集. 2008:72-79.

[2] 刘桂英,李祖华,王彬. CMAGrid 中作业调度插件的设计与实现. 高性能计算技术,2009(2):48-52.

[3] 曹燕,王彬,李娟. 国家气象应用网格平台用户安全的设计和实现//国家气象信息中心 2008 年度科技年会论文集. 2009:61-67.

[4] UNICORE Project Homepage. http://www.unicore.eu.

[5] 王彬,宗翔. UNICORE 技术调研分析报告//国家气象信息中心 2007 年度科技年会论文集. 2008:91-97.

[6] ARCON Client Library. http://sourceforge.net/project/showfiles.php? group_id=102081&package_id=127938.

[7] 刘二年,丰江帆,张宏. 基于 Flex 的环保 WebGIS 研究. 测绘与空间地理信息,2006,29(2):26-28.

[8] 王彬,宗翔,魏敏. 一个精细粒度实时计算资源管理系统. 应用气象学报,2008,19(4):507-511.

[9] 宗翔,王彬. 国家级气象高性能计算机管理与应用网络平台设计. 应用气象学报,2006,17(5):629-634.

[10] 李集明,沈文海,王国复. 气象信息共享平台及其关键技术研究. 应用气象学报,2006,17(5):621-628.

[11] 王彬. 国家气象网络计算应用节点门户系统的设计与实现. 气象科技,2006,34(增刊):5-9.

[12] 常飚. 存储检索系统监视信息采集技术分析. 气象科技,2006,34(增刊):31-35.

[13] 王彬. 一个计算作业网格执行环境的分析、设计与应用. 计算机应用研究,2008,25(8):2546-2549.

[14] 金之雁,王鼎兴. 一种在异构系统中实现负载平衡的方法. 应用气象学报,2003,14(4):410-418.

[15] Foster I. The Grid: A new infrastructure for 21st century science. *Physics Today*, 2002, 55(2):42-47.

[16] Foster I, Kesselman C, Tuecke S. The anatomy of the Grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 2001, 15(3):200-222.

Design and Implementation of Resource Monitor Module in Meteorological Computational Grid Platform

Wang Bin Chang Biao Zhu Jiang Liu Chunhua

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

Abstract

Computing resources, aggregated by meteorological computational Grid, are composed of high performance computers and storage resources. These resources are installed in different areas with different system structures, running conditions and workloads. In order to monitor the status of resources in meteorological computational Grid and to provide users and administrators with reference information, resource monitor module is designed and implemented as part of meteorological computational Grid platform software system.

The resource monitor module involves 3 layers: remote meteorological Grid nodes, resource state information acquisition, and web representation. Resource state describes the system information of high performance computers in a meteorological Grid node, comprising 3 major parts, overall information, nodes information and jobs information. The layer of resource state information acquisition is made up of poller, collector, feeder and related configuration files. Correspondingly, the acquisition process of resource state information in the resource monitor module can be divided into 3 parts, polling, collecting and feeding. Web representation layer is on the top and provides users with resource state information through commonly used internet browsers.

The resource monitor module is developed based on Grid management software UNICORE and client software ARCON Client, and implemented with Java and XML technology. ARCON Client Toolkit is used to implement node accessing function in the resource monitor module. The poller submits querying jobs for status information to computers in Grid nodes automatically and termly, and pushes it into the log queue when a job is submitted. The collector reads the queue and retrieves results of query. The feeder parses the results and writes a specially formatted XML file. The code of querying and retrieving is asynchronous so as to avoid waiting in querying. As a result, the monitor program runs stably and robustly. Major packages of resource state information acquisition layer are base driver, job scheduling, log queue query, resource state parsing, and configuration setting etc. The web representation reads the XML file containing the resource state query results, and implements resource state displaying via Flex and J2EE technologies.

At present, 10 high performance computers have been brought into centralized monitoring in National Meteorological Information Center, Beijing, Chengdu, Guangzhou, Shenyang Regional Centers as well as Anhui Province. Resource monitor module is one of the key parts of meteorological Grid platform software system and providing real time services. In the future, with the further construction of meteorological computational Grid, the resource monitor module will see further application and put major computing resources in meteorological department into supervision.

Key words: resource monitor; meteorological computational Grid; Grid platform software system; resource GIS view; computing resource