

神威集合预报可视化实时监控系统在 神威机上的设计与实现*

于 枫

(北方计算中心,北京 100091)

提 要

根据神威集合数值天气预报的运行特点,针对其运行过程中可能遇到的问题,介绍了自主研制的可视化实时监控系统的解决方案。

关键词: 实时监控系统 可视化 集合预报

基于国产神威巨型机开发的并行化集合数值天气预报系统具有以下几个特点:一是“大”,即它的解算规模大,目前该集合数值天气预报系统尚不能在国内其它机器上运行如此大的规模。二是“多”,样本个数多,达到 32 个样本。运算环节多,它由前处理、资料同化、集合预报初始场生成、模式预报、模式后处理、产品生成等 6 个子系统组成,每个子系统又包含多个步骤。使用的系统资源多,目前它使用了神威机 392 个 PE(处理器)中的 264 个,全部 16 个 IOP(磁盘阵列),存储总量达到 120 G Bytes。三是“长”,首先它运行的时间较长,4 个时次的总运行时间为 7.5 h 以内,由于受数据接收的限制,4 个时次分为两次运行,其间隔达到 2 h。

面对这样一个规模巨大、结构复杂、环节众多、样本数量大、系统资源占用多的庞大系统,我们必须要有个监控系统,才能及时处理并记录运行中出现的问题,保证它的正常运行。为了使集合数值天气预报系统更加符合业务实时、不间断运行的要求,对系统运行实行有效的控制与管理,我们使用 UNIX 环境下的 OSF/ Motif 工具库结合 C 语言,研制开发了集合预报可视化监控系统。

1 系统结构与主要功能

1.1 系统结构

集合数值天气预报监控系统对集合数值天气预报系统的解算全过程实施全面、实时、有效的监控。整个集合数值天气预报系统工作从它开始,并由它启动集合数值天气预报系统的解算,与全部解算并行运行。在解算运行过程中它始终贯穿其中,不断收集软件和

* 本文由国家“863”高科技研究发展计划(863-306-ZD11-03-02)项目资助。

2001-04-28 收到,2001-07-16 收到修改稿。

硬件两方面的信息,根据得到的信息做出相关处理,并做出详细的日志记录。在全部解算结束后,由它做出善后处理,结束整个集合数值天气预报系统的运行。

集合数值天气预报监控系统所采用的显示模型和窗口模型为 X 窗口系统,以 OSF/ Motif 图形用户界面系统 (GUI) 开发,应用程序接口为 Xlib、Xt Intrinsic、Motif Tool Kit 等,编程语言为 C 语言,其层次结构如图 1^[1]。其中用户模型、窗口模型、显示模型为可编程模型,其余三个部分为系统提供的资源。

桌面管理系统
用户模型
窗口模型
显示模型
操作系统
硬件平台

图 1 监控系统层次结构图

集合数值天气预报系统每天对 00:00、06:00、12:00、18:00 UTC 的 4 个时次的资料做滚动处理^[2],在 12:00 UTC 做 32 个样本的模式预报并得到最终的预报产品。因该系统占用资源多,运行时间长,为便于集中管理,每天的运行分为两次定时启动完成。第一次启动后依序完成对 18:00、00:00、06:00 UTC 资料的滚动处理并进行监控,这时的滚动处理只包括前处理和资料同化两部分。第二次启动后,除完成 12:00 UTC 的前处理和资料同化外,还要完成集合预报初始场生成、模式预报、模式后处理、产品生成,并对这六个部分进行监控。总体监控流程示意图见图 2。

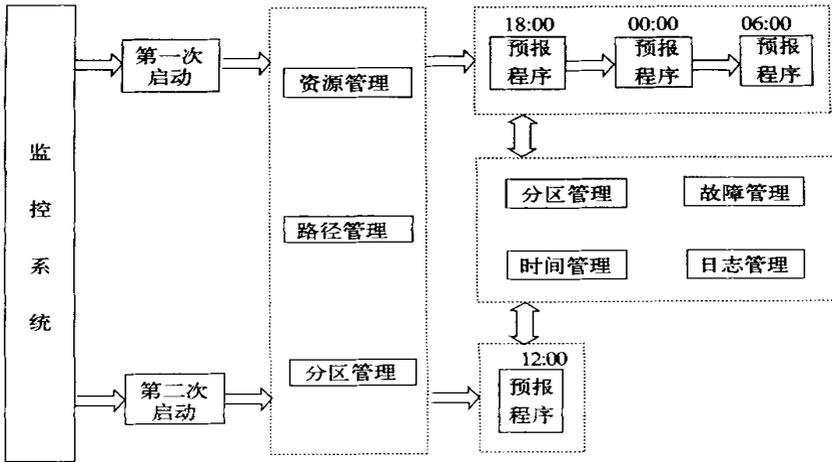


图 2 总体监控流程示意图

其中资源管理、路径管理、分区管理、故障管理、时间管理、日志管理为监控系统的六项功能。监控系统每次启动后首先执行资源管理、路径管理、分区管理三项功能,检测集合预报系统的运行条件是否满足,在运行条件满足的情况下调用各时次的预报程序。监控系统伴随预报程序执行过程执行分区管理、故障管理、时间管理、日志管理四项功能,对集合预报的解算过程进行监控。

集合数值天气预报系统包含六个子系统,每个子系统又分别有若干步(其中前处理 5 步,资料同化 2 步,集合预报初始场生成 4 步,模式预报 1 步,模式后处理 3 步,产品生成 2

步),为方便监控程序运行,达到及时了解进度的目的,在子系统内部是按照每步来进行监控,每个单步的监控结构基本相同,其监控结构如图 3。

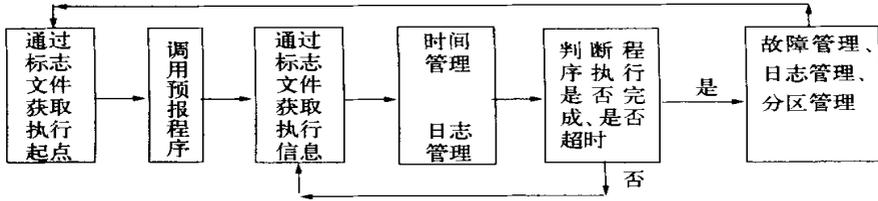


图 3 单步监控结构图

监控系统对集合数值天气预报系统的六个子系统的的所有步分别编号,通过该编号监控系统可以从任意步开始执行。预报程序启动后,监控系统通过标志文件获取预报程序提交成功的信息后对预报程序的执行时间和执行过程进行监控,同时通过标志文件获得预报程序执行的一些情况并在日志文件中进行记录。在预报程序完成或超时时,监控系统针对运行中记录的信息分别进行处理,对出现的问题做出相应的处理后将处理情况记录在日志文件中,随后根据本步完成情况决定是否可以进行下一步的运行。

集合数值天气预报监控系统的启动有两种模式:自动模式和人工模式。在自动模式时,通过在 UNIX 系统下的设置,可在设定的时间启动监控系统,监控系统自动完成系统资源检测、路径生成、分区控制后,启动执行对应时次的集合数值天气预报,并完成相应的监控操作。自动模式是集合数值天气预报监控系统的主要模式,应用于日常的业务运行工作;当因为意外原因,自动模式无法启动,或需要在非约定时间启动集合数值天气预报系统时,可采用人工模式。人工模式是由操作人员启动监控系统,在一个具有多个功能键的界面下,操作人员根据需要执行的预报程序情况,设定标志文件的初始值(初始值可以依照操作手册中的对照表来设定)并人工选择一些功能后,再提交执行集合数值天气预报的一种方式。人工模式是集合数值天气预报监控系统的辅助模式。

集合数值天气预报监控系统的内部消息传递是通过标志文件来实现的。标志文件分为两类:一类用于控制集合数值天气预报的执行,主要包括三个文件,一个时次文件,一个过程文件,还有一个产品控制文件。通过这三个文件的设置,可以控制提交的时次,提交的步骤,以及是否生成产品,并可以判断每步程序执行是否正确。还有一些文件用于记录预报程序执行过程中输出的信息,通过这些信息可以了解到预报程序执行的一些情况。另一类用于控制功能键的使用,这类标志文件主要用于人工模式,主要包括两个文件,一个控制作业是否可以提交,一个控制全部的功能键是否可以使用。通过这些标志文件的使用,操作人员可以很方便地使用监控系统的各项功能,灵活执行集合数值天气预报,同时通过人工修改时间和过程标志文件,来达到从任何起点都能提交作业的目的,从而实现了集合数值天气预报系统与全球中期业务系统 T106L19 的无缝连接。

1.2 系统功能

集合数值天气预报监控系统的结构复杂,环节众多,它不但涉及作业分区、JOP 等系统,而且要管理数据和日志信息,还要帮助运行管理人员处理解算过程中出现的问题,这

就决定了集合数值天气预报监控系统必须具有强大而全面的功能,才能满足集合数值天气预报系统实时、不间断的业务运行的需求。集合数值天气预报监控系统主要包括以下六项功能:资源管理、分区管理、路径管理、故障管理、时间管理、日志管理。

(1) 资源管理 集合数值天气预报系统使用 264 个 PE,16 个 IOP,存储总量达到 120 G Bytes,因此检测资源的状态,是否达到集合数值天气预报系统所使用的最低资源量,是保证集合数值天气预报系统正常执行的前提。资源管理主要检测 16 个 IOP 的状态是否正常,集合数值天气预报系统所需要的磁盘空间是否足够,空闲 PE 的数量是否达到 264 个。并根据检测的结果返回相应的信息,由监控系统做出相应的操作,特别是当在集合数值天气预报系统启动前发现资源缺乏时,将及时做出报告,提示管理人员准备足够的系统资源。

(2) 分区管理 集合数值天气预报系统使用两个 132 个 PE 的分区,保证这两个分区状态的正常是集合数值天气预报系统正常执行的基础。首先,监控系统具有重名检查,在集合数值天气预报系统启动时检测,防止重名分区的出现。当发现重名分区时,由于集合数值天气预报系统有很高的运行权限,因此监控系统有权自动将其它重名分区卸载,并按照集合数值天气预报系统的要求重新申请分区。其次,在集合数值天气预报系统执行过程中,监控系统不断检测分区的状态,根据检测的结果和时间管理中的超时设置来判断分区的状态。对于出现异常情况的分区,系统自动卸载作业,同时将异常分区释放,然后重新申请分区,并将作业重新提交。最后是辅助分区申请,这主要应用于产品生成,当在执行产品生成时 PE 或 IOP 出错,导致产品不能生成时,监控系统根据出错的情况可自动申请另外两个分区供产品生成程序使用,从而保证产品的生成。

(3) 路径管理 路径管理主要包括以下几项功能:一是删除磁盘上旧的数据,及时更新数据,保证新数据的及时有效;二是根据 IOP 当前状态,决定主样本的位置(监控系统将主样本放在检测到的第一个未出现异常情况的 IOP 上,同时告知主样本位于的 IOP 号,从而保证主样本能够正确执行);三是在集合数值天气预报系统启动时根据 IOP 的情况,自动生成路径控制文件(路径控制文件是程序执行过程中所需要的文件,它指明了各部分程序执行时的输入输出文件及工作文件的路径,以及生成结果的位置)。

(4) 时间管理 首先通过时间表的设定,为各程序步指定一个执行的时间上限,并以此作为判断集合数值天气预报系统运行是否正常的依据。对于那些超过时间上限而未能正常退出的作业,监控系统将根据标志文件的情况决定后续执行的步骤;其次在程序执行过程中,运行管理人员可以利用“时间管理”来查询当前程序已经运行的时间和还需要运行的时间。

(5) 日志管理 日志管理包括对预报程序执行过程中产生的屏幕输出文件和监控系统本身日志文件的管理。当集合数值天气预报系统发生故障时,监控系统将屏幕输出文件移动到指定的目录下,便于操作人员依据这些文件做出判断。监控系统的日志文件是记录每日监控系统执行情况的文件,包括了对系统资源检测,分区检测,预报程序执行情况的记录,一般保留 7 天的内容,为操作人员提供近期系统执行的情况。

(6) 故障管理 故障管理是监控系统中的重要功能,它融合在上述五项管理功能之中,针对各管理功能发现的故障情况,做出不同的处理,从而保证集合数值天气预报

系统尽可能的顺利执行。集合数值天气预报系统运行时常见的三种故障是:IOP 故障、PE 故障和数据故障。当 IOP 出现故障时,监控系统记录出故障的 IOP,如果故障发生在分区申请时,则监控系统通知分区管理,申请分区时,将放弃故障 IOP。如果在申请分区后,系统运行过程中发生 IOP 故障,则监控系统进行相应的容错处理后使系统继续运行,割断故障 IOP 及其相关的样本,由剩余样本解算而得到最终产品;当 PE 出现故障时,监控系统会做出记录并通知分区管理做出相应的操作;当数据出现故障时,集合数值天气预报系统一般无法正常执行,监控系统会弹出报警窗而给出提示,为运行管理人员提供处理依据。无论何种故障,当集合数值天气预报系统无法正常执行时,监控系统会自动将屏幕输出文件移动到指定的目录下,业务运行管理人员可根据这些文件和日志文件来判断故障的情况。

2 结束语

可视化监控系统的建立是基于运行管理人员在集合数值天气预报系统业务运行中的实际要求,针对“神威”机在运行过程中的具体状况,其功能的设置来源于运行管理人员长期工作实践经验的积累,符合集合数值天气预报系统业务运行的实时要求,可以辅助解决运行管理人员在日常管理集合数值天气预报系统中所遇到的问题。因此可视化监控系统的建立完善了整个集合数值天气预报系统的运行,进一步加强了整个系统的管理工作,大大增强了集合数值天气预报系统在业务运行中的方便性,提供了运行时的详细信息,为集合数值天气预报系统的稳定运行提供了保障,同时也极大地方便了运行管理人员的工作,为集合数值天气预报系统进入正式业务运行做出了贡献。

参 考 文 献

- 1 张倪,莫斌,等编著. Motif 与图形用户界面开发工具. 北京:清华大学出版社,1995.
- 2 国家气象中心编译. 资料同化和中期数值预报. 北京:气象出版社,1991.

THE DESIGN AND IMPLEMENTATION OF SW VISUAL REAL-TIME MONITORING CONTROL SYSTEM

Yu Feng

(North Computation Center, Beijing 100091)

Abstract

The visual real-time monitoring control scheme with error accommodation function in accordance with the characteristics and possible problems in running the SW parallel ensemble numerical weather prediction system is presented.

Key words: Real-time monitoring control system Visualization Ensemble weather forecast