

刘媛媛, 琚玲, 何文春. 自动气象站资料动态多进程处理模型的设计与应用[J]. 气象, 2012, 38(12): 1560-1564.

# 自动气象站资料动态多进程处理模型的设计与应用<sup>\*</sup>

刘媛媛 琚玲 何文春

国家气象信息中心, 北京 100081

**提 要:** 针对目前气象预报及资料同化等业务对数据服务时效的需求, 设计给出一种动态多进程的调度模型。该模型根据配置文件确定进程处理自动站资料的范围, 同时采用带参数进程对重点区域进行处理。应用结果表明, 动态多进程调度模型能够简单方便地根据应急保障需求调整自动站资料的处理策略, 以保障具有最高时效要求的资料的处理性能。通过在应对甘肃舟曲泥石流灾害等三次应用证明能使资料处理时效提高 80% 左右。

**关键词:** 自动站, 处理, 时效

## Design and Application of Dynamic Multi-Process Model for AWS Data Processing

LIU Yuanyuan JU Ling HE Wenchun

National Meteorological Information Centre, Beijing 100081

**Abstract:** Because large quantities of automatic weather station (AWS) data are increasing dramatically as well as the weather forecast system and data assimilation system are developing continually, the current processing efficiency of AWS data is unable to meet the requirement of data users. Aimed at efficiency problem of AWS data, a dynamic multi-process data processing model is proposed. In this model, configuration files are used to define handling scope of each process, and a process with parameter is designed to deal with the data of particular region which needs to be got efficiently. The model has been implemented in three cases such as disastrous debris flow at Zhouqu, Gansu Province on 7 August 2010. The result shows that data processing strategy can be adjusted easily in terms of important weather phenomena and data processing timeliness can be guaranteed, increasing data processing efficiency by 80%.

**Key words:** automatic weather station (AWS), data processing, efficiency

## 引 言

地基、空基和天基一体化的综合观测系统是现代观测体系建设的目标, 自动气象站是地基观测的重要组成部分。同时自动气象站资料是描述中尺度、短时效天气现象的观测资料, 对于分析相应天气现象发挥着非常重要的作用, 近几年随着气象预报及资料同化技术的不断发展, 对于自动站资料的服务时效也提出更高的要求<sup>[1-5]</sup>。但由于自动站资料时空分辨率的不断提高(截至 2011 年 4 月, 全国参与考核的有 30136 个自动气象站, 其中国家站 2413

个, 区域站 27723 个。自动气象站资料观测频率达到 5 分钟), 用于服务的数据存储量级别逐渐向海量数据发展, 因此如何及时快速处理海量数据成为决定数据服务时效的一个重要环节。

在国家气象信息中心, 实时数据库系统(以下简称实时库)承担着自动气象站资料的数据处理、存储管理及数据服务的功能。由于实时库建成时面对的资料量并不大, 它对于所有资料都采用单进程顺序处理的方法。但随着自动气象站报文数量的逐年上升, 这种处理方式逐渐不能满足用户对于数据处理时效的需求。而且当重大灾害发生时, 用户对数据的时效要求会进一步提高。在近几年极端天气

\* 中国气象局“区域自动站实时资料质量控制业务化推广应用”项目资助

2011 年 10 月 7 日收稿; 2012 年 7 月 18 日收修定稿

第一作者: 刘媛媛, 主要从事系统开发、数据库维护工作。Email: liuyy@cma.gov.cn

频发的情况下,实时库亟需具备更强的数据处理能力和更快的响应时间,以做好防灾减灾的应急保障。

目前,类似于实时库单进程顺序处理等传统的串行方式其在一个 CPU 指令周期内只能执行一条指令,不可能依赖无限地提高处理器的速度来提高问题的求解速度<sup>[6]</sup>。因此,为了满足大规模数据计算的需求,并行计算成为了一条有效解决问题的途径,并且目前普遍采用的也是利用并行处理技术<sup>[7-8]</sup>。

本文借鉴并行处理技术提出自动站气象资料动态多进程处理模型,该模型改变了实时库原有的静态顺序处理模式,采用动态多进程方法构造灵活的处理模型。

### 1 自动气象站资料动态多进程处理模型的设计

自动气象站资料动态多进程处理模型是在应对自动气象站资料不断增长(如图 1)的过程中逐步形成的,首先由最早的单进程顺序处理模型调整为静态多进程处理模型,然后由静态多进程处理模型调整为动态多进程处理模型。

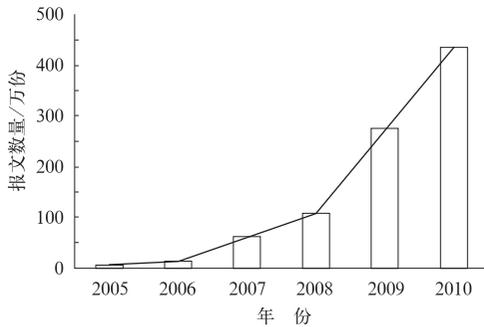


图 1 实时库处理自动站资料报文数量变化曲线

Fig. 1 Quantitative curve of AWS observation report from 2005 to 2010

### 1.1 单进程顺序处理模型调整为静态多进程处理模型

经过分析,实时库原有的单进程顺序调度处理模式是自动站地面资料处理时效无法提高的主要原因。2008 年,实时库改变原有自动站资料处理方式,由单进程顺序处理改为多进程并发处理。多进程并发并不能直接并发启动多个进程处理资料,需要解决多进程之间的通信问题,例如文件并发访问等。

实时库自动站资料静态多进程处理模型通过配置文件控制每个进程所处理的资料范围,使得每个进程各司其职,不会发生冲突。具体模型结构见图 2。

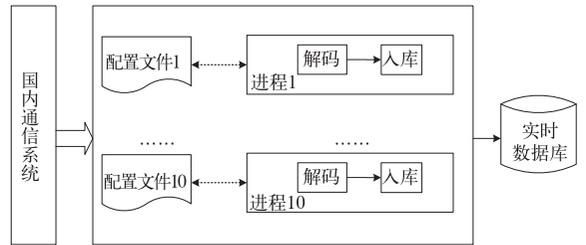


图 2 静态多进程处理模型

Fig. 2 Static multi-process model

对于如何确定每个进程处理资料的范围,实时库综合考虑用户需求及进程之间的负载均衡。自动气象站分为国家站和区域站,一般情况下国家站资料数据质量优于区域站资料,国家站资料对于用户更为重要,需要单独予以保障。因此,为满足自动站资料整体处理需求,实时库系统采用 10 个静态进程同时处理自动站资料,mrdbp\_aut1 处理单站上传的自动站地面资料,mrdbp\_aut2 处理其他类型的自动站资料,mrdbp\_aut10 专用于处理打包上传的国家级自动站地面资料。其他 7 个进程均处理打包上传的区域自动站地面资料,这 7 个进程各自处理资料的范围是基于区域自动站的地理分布的疏密情况分配的。详见表 1。

表 1 10 个进程处理资料范围

Table 1 Data range of ten processes

进程名	进程处理资料范围(省市名称)	2008 年进程处理的台站数	2010 年进程处理的台站数
mrdbp_aut1	自动气象资料单站上传的地面资料	—	—
mrdbp_aut2	自动气象站其他类型数据(如日数据,辐射资料,单要素雨量站资料,土壤资料等)	—	—
mrdbp_aut3	区域自动站地面资料之北京、内蒙古、河北、天津、山西	2817	3139
mrdbp_aut4	区域自动站地面资料之广东、广西、海南	2070	2551
mrdbp_aut5	区域自动站地面资料之安徽,江西,山东	3482	4019
mrdbp_aut6	区域自动站地面资料之上海,江苏,浙江,福建	2443	2932
mrdbp_aut7	区域自动站地面资料之湖北,河南	2889	3031
mrdbp_aut8	区域自动站地面资料之湖南	2107	2127
mrdbp_aut9	区域自动站地面资料之四川、重庆、贵州、西藏、吉林、黑龙江、甘肃、陕西、青海、宁夏、新疆、云南、辽宁	5841	7920
mrdbp_aut10	国家站地面资料	2135	2416

为了增加程序的灵活性,每个区域站地面资料进程所处理的资料范围都是通过外部配置文件确定的。以 mrdbp\_aut3 进程的配置文件的配置样例:

外部配置文件名: aut3\_type. cfg

内部格式:

CCCC 数量 具体 CCCC(以空格区分)

5 BEPK BEHT BESZ BETJ BETY

进程配置参数化设计为后续灵活动态调整提供了基础。

### 1.3 静态多进程处理模型调整为动态多进程处理模型

从图 1 和表 1 均可以看出,相比较 2008 年,2010 年的自动站数量又有大幅增加,虽然采取了静态多进程的处理方式,自动气象站资料时效问题获得基本解决,但还是无法满足用户更高层次的需求。在 2010 年甘肃舟曲泥石流灾害的应急保障过程中,用户需要及时了解当地实况,但事实上灾区部分站点数据有时需要延时 30 分钟左右用户才能得到,这种时效远远不能达到应急保障的要求。实时库需要实现能够对于重大灾害天气影响地区的观测资料优先处理并提供服务,但原有的静态多进程处理模型并未考虑到这方面的需求。

为提高实时库应急保障水平,实时库设计了动态多进程处理模型。该模型在静态多进程模型的基础上,增加了一个动态程序 mrdbp\_aut0。该程序可以通过设置不同参数启动多个进程处理气象灾害影

响区域的观测资料。该程序具有灵活的区域参数配置,区域的参数设定为影响区域文件名中的编报中心(CCCC)关键字,这样可以达到根据需要进行动态调整的目的,能够为用户急需使用的资料建立绿色通道,提高应急保障地区资料的处理时效。正常模式下,动态多进程处理模型就是原静态多进程处理模型。在应急保障模式下,该模型可增加处理进程,动态地将应急保障区域设置为进程参数进行运行。

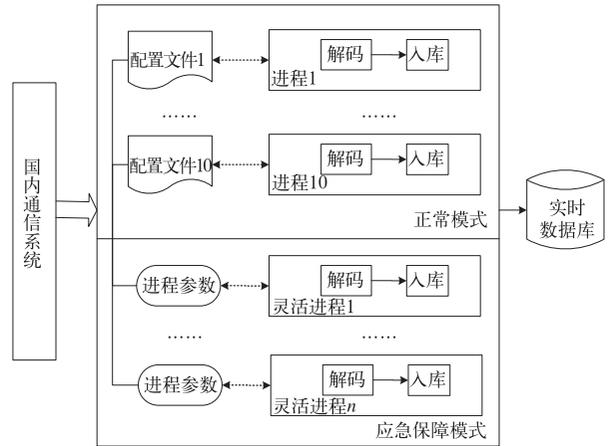


图 3 动态多进程调度模型

Fig. 3 Dynamic multi-process model

使用动态多进程调度模型能够改进静态多进程模型的存在的一些不足,使得文件调度处理方法适合于应急保障需求。如, mrdbp\_aut9 进程有三个待处理文件,见表 2。

表 2 三个待处理文件情况  
Table 2 Information of three files

文件编号	文件名	报文内容	报文优先级
1	Z_SURF_C_BEHB-REG_20101206001358_O_AWS_FTM. txt	黑龙江省 00 时 10 分 部分报文	2
2	Z_SURF_C_BEYC-REG_20101206001824_O_AWS_FTM. TXT	宁夏自治区 00 时 10 分 部分报文	2
3	Z_SURF_C_BCLZ-REG_20101206000604_O_AWS_FTM. TXT	甘肃省 00 时 部分报文	1

文件 3 为重大灾害发生时受灾区域的观测资料,用户重点关注该区域整点时次的观测资料,所以文件 3 处理的优先级要明显高于文件 1 和文件 2。但在实际的静态多进程模型下,由于系统处理软件采用的读取目录文件 readdir 函数在 HP-UNIX 系统中对目录文件的排序规则存在不确定性,这三个文件的调度顺序会出现图 4 的情况,即先到达的受灾区域观测资料反而不能得到及时地处理。从文件

开始处理时间来看,文件 3 最先收到(00:08:38),却最后被处理(00:23:35),处理时间比文件收到时间推后了 15 分钟,严重影响受灾区域观测资料的入库时效。

根据动态多进程处理模型的设计思路, mrdbp\_aut0 进程会专门处理受灾地区的资料,从而使得灾区资料得以快速处理。如采用该模型,理论上这三个文件调度顺序如图 5。

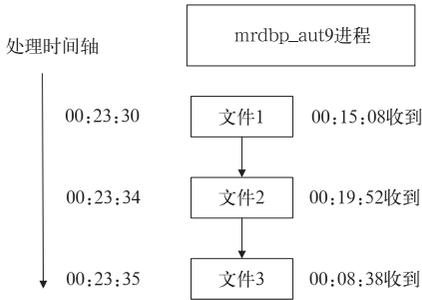


图 4 静态多进程模型调度顺序  
Fig. 4 Scheduling sequence in static multi-process model

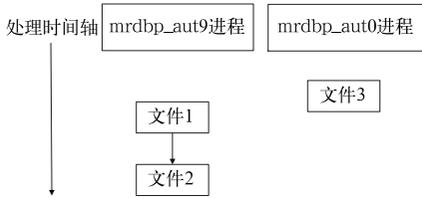


图 5 动态多进程模型调度顺序  
Fig. 5 Scheduling sequence in dynamic multi-process model

## 2 应用效果

### 2.1 在应对舟曲泥石流灾害的应用

2010年8月7日甘肃舟曲特大泥石流灾害,8月8日中国气象局启动重大气象保障Ⅱ级应急响应命令,要求各部门全力做好此次灾害气象保障。为了更高效地为用户提供受灾地区的实况数据,实时数据库系统调整数据处理方法,采用动态并发的处理模型,同时将用户关注的甘肃、四川、陕西三省设置为重点区域,对于这三个省的资料采用 mrdbp\_aut0 进程进行特殊处理。

图6为使用动态多进程处理模型调整前后三省

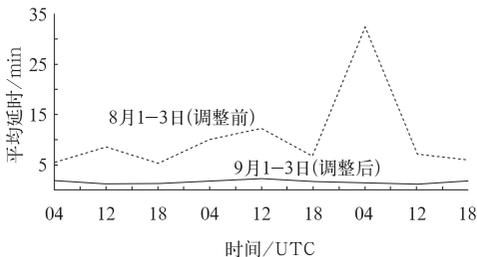


图 6 调整前后三省资料入库延时的对比情况  
Fig. 6 Comparison of processing efficiency of AWS observation data (Gansu, Sichuan and Shaanxi) between before (dashed line) and after (solid line) model adjustment

资料入库延时的对比情况,调整前三省资料平均延时为 10.4 分钟,调整后平均延时为 1.7 分钟,提高了 84%。

### 2.2 在应对“海马”台风中的应用

2011年6月22日,第04号热带风暴海马在广东登陆。受“海马”和西南季风的共同影响,广东沿海、福建南部沿海、海南中东部等地有大到暴雨,局部大暴雨,为做好应急数据服务,实时库系统将广东、海南、贵州、广西、福建设置为重点区域,由于广东、贵州在之前已经被设置为重点区域,海南、广西两省原本就在一个进程(mrdbp\_aut5)中,而且该进程只处理这两个省的资料,因此这次只将福建省增加为重点区域。图7为使用动态多进程处理模型调整前后福建省资料入库延时的对比情况,调整前福建省资料平均延时为 9.4 分钟,调整后平均延时为 1 分钟,提高了 89%。可以看出调整前福建省资料入库延时很不稳定,而调整后资料入库延时基本稳定保持在 1 分钟左右。

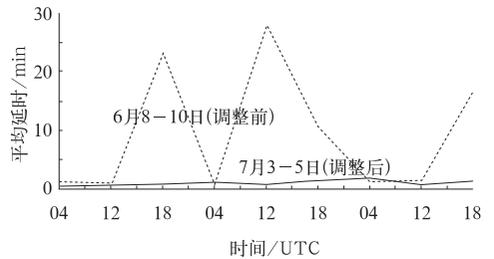


图 7 调整前后福建省资料入库延时的对比情况

Fig. 7 Comparison of processing efficiency of AWS observation data (Fujian) between before (dashed line) and after (solid line) model adjustment

### 2.3 在不同硬件环境服务器中的应用

前两例应用都是在 HP superdome 服务器(实时库业务系统)上运行。为了考察动态多进程处理模型在不同硬件环境服务器的运行情况,在应对“南玛都”台风时,计划在性能更好的 IBM P570 服务器(实时库升级系统)上对该模型进行应用。

2011年8月31日,第11号热带风暴南玛都的中心在福建省晋江市沿海登陆,受“南玛都”影响,福建、浙江等地出现暴雨。这次将福建省、浙江省增设置为重点区域。图8为使用动态多进程处理模型调

整前后福建省、浙江省资料平均入库延时的对比情况,调整前资料平均延时为 1.36 分钟,调整后平均延时为 0.24 分钟,提高了 83%。可以看出,在实时库升级系统中自动气象站地面资料的处理性能有了整体的提高,但应用动态多进程处理模型仍能进一步提升重点区域的入库时效。

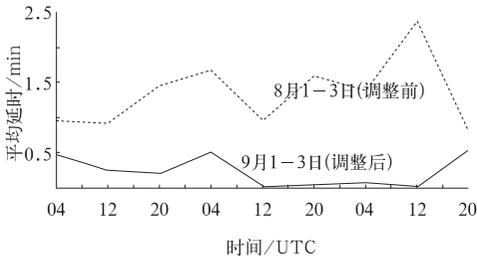


图 8 调整前后福建省、浙江省资料入库延时的对比情况

Fig. 8 Comparison of processing efficiency of AWS observation data(Fujian and Zhejiang) between before (dashed line) and after (solid line) model adjustment

### 3 结 语

资料处理时效决定着实时库对用户提供的服务质量,本文设计给出了采用多进程并发机制与动态

进程相结合的灵活处理模型,即动态多进程处理模型。然后将模型在实际甘肃舟曲泥石流等灾害的应急保障过程中进行应用。应用结果表明,动态多进程调度模型能够简单方便地根据应急保障需求调整自动站资料的处理策略,以保障具有最高时效要求资料的处理性能,使资料处理时效提高 80%左右。同时该模型在不同硬件环境服务器上运行情况良好。

### 参考文献

- [1] 张文建. 世界气象组织综合观测系统(WIGOS)[J]. 气象, 2010,36(3):1-8.
- [2] 矫梅燕. 天气业务的现代化发展[J]. 气象,2010,36(7): 1-4.
- [3] 李湘. 气象通信系统发展与展望[J]. 气象,2010,36(7): 56-61.
- [4] 冯筠,高峰,黄新宇. 构建天地一体化的全球对地观测系统—三次国际地球观测峰会与 GEOSS 地球科学进展[J]. 2005, 20(12):127-133.
- [5] 孟昭林,李雁,陈挺,等. 综合气象观测系统业务运行综合评估技术研究[J]. 气象,2011,37(2):219-225.
- [6] 杨超,徐如志,杨峰. 基于消息队列的多进程数据处理系统. 计算机工程与设计[J]. 2010,31(13):3128-3131.
- [7] 李惠云,殷人昆,冉望. 基于海量数据的集群服务管理模型[J]. 计算机应用,2008,28(5):1316-1318.
- [8] 周钦强. 一种利用并发提高数据处理吞吐率的模型[J]. 计算机系统应用,2010,19(11):190-194.