

杨芳,邢丽平,杨代才,等. MICAPS 历史资料存储系统设计与实现[J]. 陕西气象,2018(2):25-28.

文章编号:1006-4354(2018)02-0025-04

MICAPS 历史资料存储系统设计与实现

杨芳,邢丽平,杨代才

(湖北省气象信息与技术保障中心,武汉 430074)

摘要:为实现多种类数据的集中采集、统一管理,在省级部署 MICAPS 历史资料(经 MICAPS4 后台解码程序所生成的文件类型的数据产品)存储平台。该平台能很好地满足湖北省各级台站对 MICAPS 历史资料的需求,为 MICAPS4 在全省全面业务化做好准备工作。通过对该系统的架构进行描述,并以相关数值预报产品为例,介绍数据存储、补调、恢复流程,以期与其他省级部门相关业务人员就 MICAPS 历史资料的管理进行良好沟通和交流。

关键词:MICAPS4;历史资料;NAS;存储;数值预报云

中图分类号:TP399

文献标识码:B

作为首个 MICAPS4(MICAPS 系列最新版本)系统运行省级试点,该系统已在湖北省气象局试运行一年多时间。此系统以首次集成集合预报、格点预报等业务功能,读取卫星、雷达以及数值预报产品的效率达到毫秒级响应时间,分布式存储系统零故障等特点,实现了全省 MICAPS 客户端直连省级分布式存储系统,为湖北现代天气预报业务提供了很好的支撑。其采取的非关系型分布式数据库为气象实时数据量身打造了一套集群存储系统,其读写性能优异,从上亿个数据中检索一个数据用时仅几毫秒^[1];为确保该实时数据

库其优异的读取性能,对所存储的数据产品有严格的时间限制,如:存储最新 30 天的数值预报产品、卫星云图资料;90 天的高空地面实况资料等。

对于预报业务人员而言,MICAPS 资料是进行天气分析跟踪,天气衍变规律总结的重要数据基础,特别是关键性、灾害性、转折性天气的历史资料更是提高预报准确率的前提^[2]。尽管 MICAPS4 系统提供数据访问接口,然而对大部分基层台站预报技术人员而言,通过数据访问接口来保存历史天气个例资料依旧比较困难。因此在省级搭建一个稳定、可靠、能充分考虑资料共享与使

收稿日期:2017-12-08

作者简介:杨芳(1982—),女,汉族,湖北洪湖人,学士,高级工程师,从事气象信息通信研究。

基金项目:湖北省气象局科技发展基金重点项目(2015Z02)

[7] 栗珂,王玉玺,戴进,等. 陕南山地烤烟区小网络气温场和降水场的分析与应用[J]. 陕西气象,2001(4):17-20.

[8] 王玉玺,栗珂,韦成才,等. 陕南适宜烤烟气候条件及区域划分的研究[J]. 陕西气象,2001(5):15-18.

[9] 张振平. 适宜烤烟区划理论与实践[M]. 西安:陕西科学技术出版社,2007.

[10] 孟亚利,曹卫星,周治国,等. 基于生长过程的水稻阶段发育与物候期模拟模型[J]. 中国农业科学,2003,36(11):1362-1367.

[11] 韩锦峰. 烟草栽培生理[M]. 北京:中国农业出版社,1996.

[12] 操筠. 恩施烟草与气象条件的关系及气象灾害风险评估[D]. 武汉:华中农业大学,2014.

[13] 曲静. 调和权重法在作物产量预报中的应用[J]. 陕西农业科学,2011,57(5):109-110.

[14] 胡雪琼,朱勇,张茂松,等. 云南省烤烟农业气象灾害监测指标的研究[J]. 中国农业大学学报,2009,14(1):90-94.

用便利性的 MICAPS 历史资料存储架构环境,成为 MICAPS4 在湖北省正式业务运行前迫切需要解决的问题。

1 系统设计思路

1.1 集中采集与统一管理

系统能实时采集 CIMISS 数据库中的实况数据,接收不同系统或服务器所推送 MICAPS 原始数据,并进行实时解码处理,生成文件格式的数据产品;同时将接收到的原始数据进行自动备份,最终为用户提供一个以文件形式存储的涵盖 MICAPS 原始数据与数据产品的数据平台。省级以及省内各级台站可通过文件共享或 FTP 下载方式来访问该平台以获取所需的数据。

1.2 数据完整与准确性

系统在省级部署,可确保能为历史资料的应用提供完整准确的数据源,同时极大减轻市县两级维护系统与数据的工作量。对所备份资料进行分类^[3],每种数据采用一天一个压缩包的保存方式(逐日定时分类压缩过去 24 小时的 MICAPS 数据文件)。数据存储的目录结构定为 4 级,第 1 级目录为数据生成的年份,第 2 级目录为数据生

成的月份,第 3 级目录为资料类型,第 4 级目录为资料类型名+日期的数据文件(例:ECMWF_HR20180102.zip),清晰的目录结构便于日后对历史数据进行管理和维护。

1.3 海量存储与高扩展性

系统所存储的业务数据量主要来自于两个部分,一是解码后的数据产品文件(用户通过 MICAPS 客户端可以直接调用),至少保存 3 个月;二是备份的经压缩过的原始数据及数据产品,至少存储 1 年(表 1)。考虑到今后业务发展趋势,本系统建设初期具备 240 TB 的存储能力,支持至少 1 年的在线数据服务。

针对 MICAPS 数据产品小文件数目多、整体数据量大等特点,采用一体化的存储解决方案,选择 NAS 的存储架构。NAS(网络附加存储),即将存储设备通过网络连接到现有的服务器上(用户通过访问服务器来获取 NAS 里的数据文件),基于 TCP/IP 协议提供数据和文件级数据的存取服务。在支持大容量的同时,也具备大规模横向扩展能力,即在不做数据迁移的前提下,实现存储容量和性能的平滑扩展。

表 1 存储的资料类型与在线存储时间

序号	资料类型	在线保存时效	
		数据产品	原始数据
1	T639 高分辨率模式产品(T639)		
2	欧洲中心高低分辨率数值预报产品(ECMWF_HR)		
3	欧洲中心低分辨率数值预报产品(ECMWF_LR)		
4	日本高分辨率数值预报产品(JAPAN_HR)		
5	德国高分辨率数值预报产品(GERMAN_HR)	不保存	
6	中尺度数值预报产品(GRAPES_MESO_HR)		1 年以上
7	全球中期数值预报产品(GRAPES_GFS)		
8	NCEP 模式产品(NCEP_GFS)		
9	华东模式产品(SHANGHAI_HR)		
10	欧洲集合预报(ECMWF_ENSEMBLE)	—	
11	高空站点资料(UPPER_AIR)		—
12	地面站点资料(SURFACE)	1 年以上	
13	卫星资料(FY2E、FY2F、FY2G)、日本葵花 8 号卫星资料		不保存

注:表中数据产品与原始数据是指将数据经压缩程序处理过的压缩包;“—”表示无此类数据文件

1.4 备份策略与资料类型

解码后的数据产品文件,直接存放在 NAS 指定目录下(不列入下表 1)。对需要压缩备份的 MICAPS 原始数据与产品,可根据数据量大小、目录结构、业务需求的不同,选择不同备份方式^[4];对原始数据,采取直接压缩备份方式(ZIP 压缩包);对数据产品先使用 Rsync 精确同步指定格式文件再压缩备份(TAR 包)。用于数据备份的脚本运行在 Linux 系统下,主要使用 Python 语言完成。

正常情况下,系统在每天设定的时间分别对 10 类 MICAPS 原始数据、3 类 MICAPS 数据产品前一天的资料进行自动压缩备份,具体信息见表 1;同时部署数据监控脚本,当自动备份不成功时,可以运行手动备份脚本完成相关数据文件的备份。

2 系统流程与实例

2.1 系统流程

该系统由数据处理服务器(Red Hat Enterprise Linux 7.1 操作系统)、NAS 网络附加存储两部分组成(Linux 下 NFS 挂载 NAS)(见图 1),系统能实时处理由 CTS 或省级文件服务器推送的 MICAPS 原始数据,同时读取 CIMISS 数据库中的地面、高空实况数据并进行解码处理,最后生成以文件形式存储的 MICAPS 格式的数据产品,保存在 NAS 指定目录中。对于原始数据与数据产品的备份^[5],则是利用 Linux 的 Crontab 功能每天定时调用资料备份脚本对相关数据进行压缩备份,存储至相应的目录(见图 2),其中实时数据流式解析系统(DPC)与站点轮询系统(JobSchedule)均来源于 CIMISS-MICAPS4 后台数据环境。

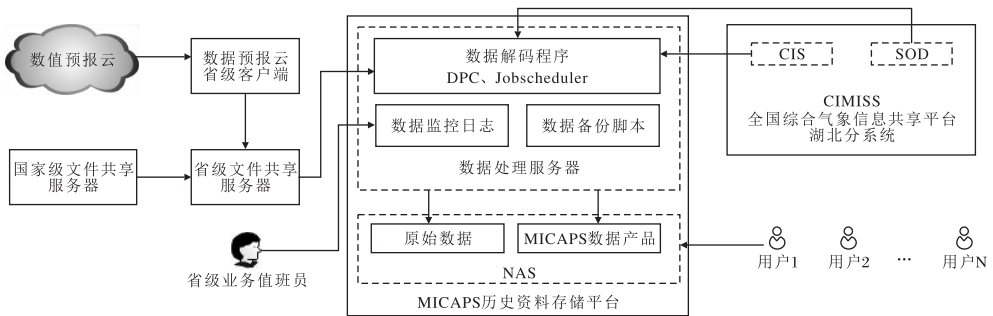


图 1 MICAPS 资料处理、存储流程图

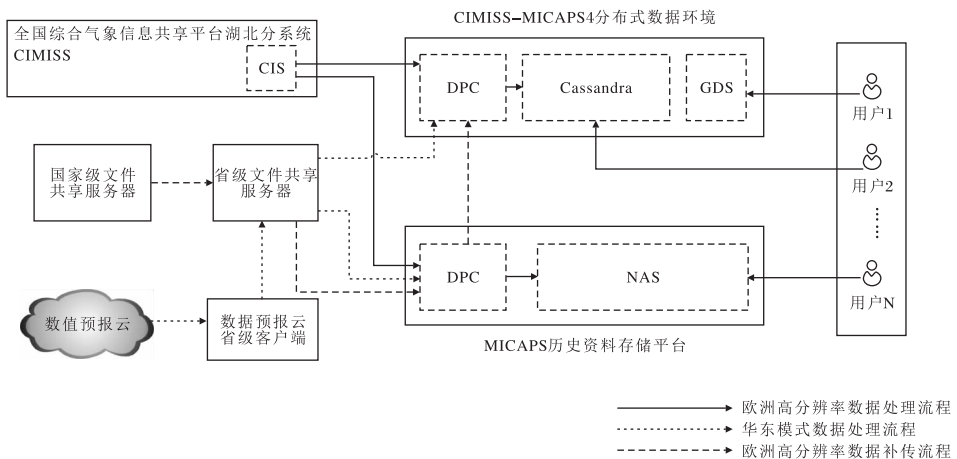


图 2 数值预报数据处理与补传流程

通过部署在该平台上的监控脚本来判定所备份数据的完整性(检测所存储数据压缩包的大小),并将告警信息写入监控日志。值班人员通过

查看监控日志,来判断是否需要补调原始数据、恢复数据产品以及手动运行备份脚本等。

湖北省各级气象台站不仅可以通过 CIMISS-

MICAPS4 分布式数据环境提供的接口 (GDS)调用 MICAPS 数据产品,也可以通过 FTP 或文件共享方式连接到该存储平台获取 3 个月以内的文件数据产品以及 1 年以上的原始数据;对于超过三个月的数据产品,可由系统管理员将备份的原始数据导入该系统后,用户再行调用。该存储系统的应用,不仅降低了省级技术人员运维压力,也确保了 MICAPS 原始数据与数据产品的完整性。

2.2 应用案例

2.2.1 华东数值预报数据

随着各地精细化预报服务需求的日益增长,各区域气象中心均建立 3~9 km 的高分辨率区域数值天气预报业务系统。为推进区域高分辨率数值预报产品的快速共享,提升区域高分辨率数值预报模式对短时临近、精细化预报等业务的支撑能力,基于 INTERNET 和公共云资源,由上海市气象局牵头建立了数值预报云,为全国各省提供华北、华东等数值预报产品实时下载等服务。以华东数值预报产品为例,梳理该数据传递给预报员并提供数据共享的流程。

省级数值预报云客户端实时下载华东模式预报产品,采用 FTP 方式共享。省级文件共享服务器从该客户端上下载华东模式产品后,FTP 推送给 MICAPS4 及 MICAPS 历史资料存储平台。两个平台在完成数据的解析后,将解码后的数据保存至 Cassandra 或 NAS 中,供全省各级台站业务人员调用(见图 2)。

2.2.2 欧洲高分辨率模式数值预报产品

与华东、华南数值预报产品由数值预报云提供相比,大部分气象模式数据由 CTS 推送。以欧洲高分辨率模式数值预报产品为例(见图 2),对该类数据共享、补传、历史数据恢复流程进行介绍。

欧洲高分辨率数据由 CTS 推送给 MICAPS4 后台数据环境及 MICAPS 历史资料存储平台。一旦该原始数据出现缺失,可由省级文件共享服务器从国家级文件共享服务器上调取原始数据至 MICAPS 历史资料存储平台,完成资料的补传,确保存储平台上原始数据与解码后数据的完整性。MICAPS 历史资料存储平台上完整的数据源,也可成为 MICAPS4 系统的另一数据来源,在

局域网环境下提升数据补调效率。

用户通过 FTP 或文件共享可获取的 MICAPS 历史资料(类型见表 1),若需要 3 个月以前的数值预报数据,则需要省级系统管理员进行历史数据恢复操作。操作流程见图 3,首先将保存在 NAS 中的原始数据压缩包解压至数据处理服务器的指定目录,再通过 DPC 控制台交互程序,手动恢复数据(可同时恢复不同种类不同时段的数据),恢复出的数据产品直接写入 NAS 数据产品的目录中,供用户调用。

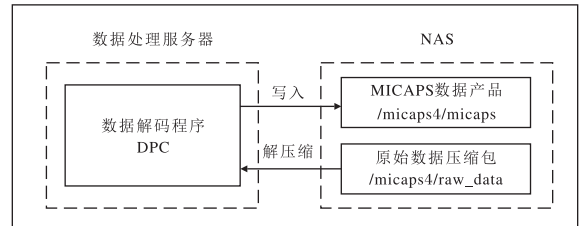


图 3 欧洲高分数值预报历史数据恢复流程

3 结语

为更好推进 MICAPS4 在湖北省的全面业务化,利用 NAS 搭建 MICAPS 历史资料存储系统,为各级气象台站提供多种 MICAPS 数据产品获取方式。试运行以来,系统稳定,实现了市县两级运维人员对该系统“零”维护。基于 NAS 的存储方式,将存储设备与服务器分离管理,减少了服务器负担,提高了存储效率。

参考文献:

- [1] 王若瞳,黄向东,张博,等.海量气象数据实时解析与存储系统的设计与实现[J].计算机工程与科学,2015,37(11):2045-2054.
- [2] 李月安,曹莉,高嵩,等.MICAPS 预报业务平台现状与发展[J].气象,2010,36(7):50-55.
- [3] 窦以文,卢俐,刘旭林,等.气象数据存储管理系统[J].计算机系统应用,2011(7):116-120
- [4] 林玉成,赵瑞,罗兵.基于 Rsync 的中央气象台数据备份机制及优化设计[J].高原山地气象研究,2014(1):81-85.
- [5] 刘斌.基于 Linux 的文件实时备份系统设计与实现[D].北京:华北电力大学,2014
- [6] MarkLutz. Learning Python[M].北京:机械工业出版社,2011.