文章编号:2096 – 5389(2022)05 – 0115 – 06

基于气象业务零中断的存储数据迁移探讨研究

谢礼江

(广东省气象探测数据中心,广东　广州　510088)

摘　要:在广东省气象局 2015 年建设高性能存储设备逐渐老化,厂商即将停止设备技术支持服务背景下,针对存储所承载的气象业务系统面临着数据丢失或业务中断的风险问题,结合气象业务时效性、连续性等特点,通过运用异构迁移技术,对新存储建设要求、数据迁移原则、存储规划、风险评估、回退方案、数据迁移进行探讨研究,实现 50 多个核心关键业务系统的 430 TB 数据在 82 h 完成平稳迁移,数据迁移速度最高达 2.1GB · s^{-1},平均速度 1.5GB · s^{-1},相比在应用层上做数据迁移的传统方法,异构虚拟化迁移技术具有迁移速度快、时间短、工作量少、业务无中断和可集中迁移的特点。

关键词:异构迁移技术;数据迁移;气象业务

中图分类号:TP333　**文献标识码**:B

Research on Storage Data Migration Based on Zero Interruption of Meteorological Service

XIE Lijiang

(Guangdong Meteorological Observation Data Center, Guangzhou, 510088)

Abstract: The high – performance storage equipment constructed by Guangdong Meteorological Bureau in 2015 is gradually aging and the manufacturer is about to stop the equipment technical support service. In view of the risk of data loss or business interruption of the meteorological business system carried by the storage, considering the timeliness and continuity of meteorological business, by using heterogeneous migration technology, the requirements of new storage construction, data migration principles, storage planning, risk assessment, falllback scheme and data migration are discussed and studied. The 430 TB data of more than 50 core critical business systems can be smoothly migrated in 82 hours, and the data migration speed is up to 2.1GB · s^{-1}, and the average speed is 1.5GB · s^{-1}. Compared with the traditional method of data migration at the application layer, heterogeneous virtualization migration technology has the characteristics of fast migration, short time, less workload, non – stop business and centralized migration.

Key words: heterogeneous migration technology; data migration; meteorological operation

0　引言

2015 年,广东省气象局建设的华为 OceanStor 18800 高性能存储投入业务运行以来,在支撑气象业务系统中发挥了非常重要的作用。这套高性能存储是广东省气象业务网、IDEA 接口平台、SWIFT[1]、基础设施虚拟化资源池平台、广东省气象决策辅助系统、广东省省突发布管理系统、FAST3.0 业务平台、CIMISS 等 50 个核心关键业务系统的存储资源支撑。随着这套存储满负载运行年限增加

收稿日期:2021 – 02 – 27

作者简介:谢礼江(1984—),男,硕士,工程师,主要从事气象信息系统研发、气象私有云虚拟化资源池服务等工作,E – mail:286139335@qq.com。

(已运行 6 a)和逐渐趋于老化,存储硬件设备进入 IT 产品生命周期(5 ~ 7 a)末段,性能处于下降的趋势,故障率也逐年提高[2]。仅在 2020 年,就更换隐患或故障盘 30 个,升级控制器软件 3 次,更换其它硬件设备 8 次,全年平均每 15 d 就需要处理 1 次故障,每月至少出现 1 ~3 个隐患盘,机框、主板、CPU、内存、网卡和 HBA 卡等硬件设备也无规律地出现故障,且此套存储硬件设备逐渐断供,厂商计划将在 2021 年 12 月 31 日停止对 OceanStor 18800 高性能存储提供技术维保服务,这将对相关气象业务的安全保障和业务连续性构成极大风险。故障会导致业务受到影响,甚至导致业务中断或数据丢失,不利于气象业务系统稳定运行。

为保持现有业务的连续性,以及满足近年增加的气象业务,急需建设一套高性能存储替换现有存储。由于气象业务具有时效性、连续性和稳定性[3]的特点,需要基于气象业务零中断的基础上,探讨研究如何把存放在 OceanStor 18800 高性能存储的数据迁移到新存储上,实现这 50 多个核心关键业务系统的平滑迁移。

1　OceanStor 18800 存储情况

OceanStor 18800 存储采用 RAID2.0 + 块虚拟化卷架构,所有磁盘阵列柜全部配置为高性能 15K 转速的 SAS(Serial Attached SCSI)磁盘和固态硬盘(SSD:Solid State Disk)[4],以高性能、高可靠、高扩展、存储效率、数据保护为其设计理念,总共配置了 6 个控制器,1TB 缓存,占用机房机柜 4 个,总可用容量 550 TB,组成多活控制器群集,为业务端应用主机访问存储提供了负载均衡和高可用功能。

OceanStor 18800 存储根据气象业务需求,目前分配 13 个硬盘域,13 个存储池。划分 20 个卷 LUN(logic unit number)约 500 TB 作为虚拟化集群资源池使用[5],运行了 700 台虚拟服务器机,部署了超 50 个核心关键气象业务系统;60 个卷约 50 TB 空间作为 CIMISS 核心业务数据库系统的 Oracle RAC 群集使用。

2　新存储建设要求

Oceanstor 18800 高性能存储承载着广东省气象台、广东省探测数据中心、广东省气象服务中心等 10 个单位的核心关键气象业务系统服务。为保持现有业务的连续性,以及满足近年增加的气象业务存储需求,在选择新存储时需满足以下几点要求:

性能和存储空间更优。气象业务飞速发展,基础设施资源池的存储资源是业务系统的底层支撑,其性能和存储是否满足业务需求直接影响着气象现代化和信息化的发展速度。

能耗和占用物理空间更小。减少耗能可以降低成本,也是人类在进行任何生产活动追求的目标;减少占用物理空间,提高机房使用率。

支持在线数据迁移。根据气象业务时效性、连续性和稳定性的特点,气象业务的安全保障和业务连续性是最优先考虑的因素,无法在线数据迁移意味着要停 50 个关键核心的业务系统,业务系统的中断会对气象业务造成很大影响。

按照新存储建设要求思路,本文选择同品牌和同系列且性能更优、扩展性更好的华为 OceanStor18810 V5 “芯”系列高端智能混合闪存存储代替旧存储。相比旧存储,OceanStor18810 V5 继承了旧存储块级虚拟化、智能缓存分区技术、同步远程复制、异步远程复制、智能数据迁移、智能数据迁移、异构虚拟化等技术特点,并针对旧存储存在的架构设计和硬盘设备老旧、占用物理空间多、耗能高等问题研发,运用 SmartMatrix 3.0 架构、闪存优化技术、SAN 与 NAS 一体化双活、高效能的硬件平台,为数据存储和使用提供可靠性更强、性能更好、业务数据迁移更稳定的解决方案。

OceanStor18810 V5 采用智能矩阵式多控架构,以控制框为单位横向扩展,达到性能和容量的线性增长,运用 4U Acitve – Active 四控冗余高密架构设计,每个控制框支持 4 个控制器和 2 个控制器 2 种方式,在提高性能的同时,又减少了占用的物理空间,相比 Oceanstor 18800 旧存储,各方面都具有很大优势(表 1)。

3　数据迁移探讨

3.1　数据迁移原则

本次存储涉及 VMware 虚拟化集群资源池和 2 台服务器 CIMISS ORACLE 数据库,资源池承载着约 700 台虚拟机,运行了超过 50 个气象关键核心系统,采用异构迁移技术在存储底层进行约 430 TB 的数据、多应用的混合模式迁移,必须基于安全性、可行性和可操作性的原则进行探讨研究,减少数据迁移过程中对气象业务造成的影响甚至数据丢失风险。

安全性。在数据迁移过程中,数据的安全放在首位,整个实施调优方案也是以数据安全为出发点进行设计。

表 1　新旧存储性能对比

Tab. 1　Performance comparison between old and new storage

指标	18800 V1(旧)	18810 V5(新)	说明
控制框	6U,六控, 1 TB 缓存	4U, 四控, 2 TB 缓存	正偏离,控制器减少,缓存增加
硬盘	1027(个)	204(个)	正偏离,减少 823 个
硬盘框	43(个)	10(个)	正偏离,减少 33 个框
占用空间	4 柜/共 129U	2 柜/共 48U	正偏离,空间占用减少 50%
性能	19 万 IOPS	26.5 万 IOPS	正偏离,性能提高 39%
SAN&NAS	否	是	正偏离,功能增加
能耗	8500 W	5070 W	正偏离,能耗降低 40%
总可用容量	550 TB	651 TB	正偏离,容量增加 18%

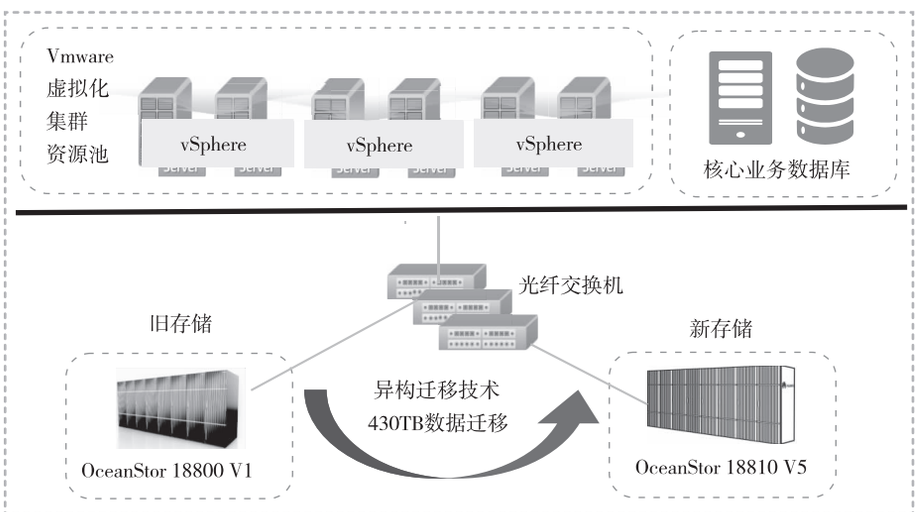

图 1　存储迁移网络图

Fig. 1　Storage migration network diagram

可行性。数据迁移综合考虑业务需求,环境情况和应用部署情况,根据收集到的信息进行充分考虑,对数据迁移和业务要求评估讨论,确保数据迁移可行。

可操作性。整套数据迁移也同时考虑在实施过程中操作的难易程度,工作量,复杂度等因素,要求可操作性强,降低气象业务影响风险系数。

3.2　数据迁移工作原理

异构虚拟化迁移技术主要是通过把异构阵列映射到本端阵列,把异构阵列的存储空间通过 eDevLUN(ExternalDevice LUN)的方式管理和利用起来。元数据卷用于对 eDevLUN 的数据存储位置进行管理,在本端存储系统上创建的 eDevLUN 与异构存储系统上的外部 LUN 形成一一对应的关系,对 eDevLUN 的读写操作实现了对外部 LUN 的数据访问。通过 LUN 伪装技术,让存储系统的 eDevLUN 的 WWN 和 Host LUN ID 设置成与异构存储系统上的 LUN 的信息一致,在数据迁移完成后,通过主机多路径软件实现在线 LUN 的无缝切换,从而在主机不中断业务的情况下完成数据迁移。

使用 MigrationDirector 存储数据迁移工具具有全自动、高效并发、灵活设置的特点,通过管理服务器自动推送迁移服务器,存储自动挂载和卸载,运用多并发执行任务从源端存储搬运到目的端存储,提高迁移速度和效率;可根据实际场景灵活配置线程数和启动时间,保障业务迁移的灵活性和弹性,避开气象业务高峰期,减少业务受影响风险。

MigrationDirector 工具通过以太网同存储和业务主机相连,使用 SSH 协议连接源存储和目的存储的 22 号端口,以及 REST 协议连接目的存储 8088 号端口,通过 SSH 协议连接 2 台业务主机执行存储命令。

3.3　数据迁移

3.3.1　存储规划　存储划分卷满足 VMware 虚拟化集群资源池和 CIMISS Oracle 数据库 2 个应用场景。为确保数据完整性、数据迁移安全性和存储性能一致性,本次新存储的硬盘域配置和旧存储的硬盘域配置保持基本一致。

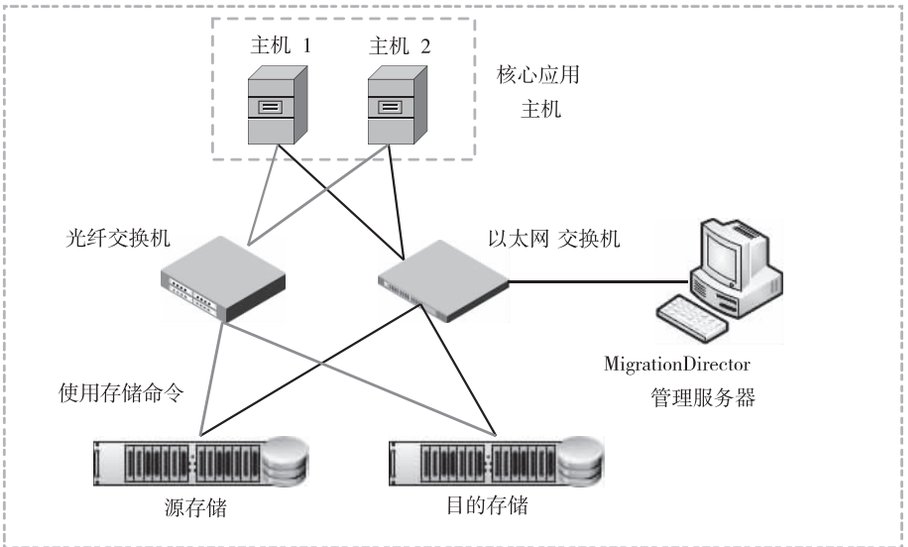

图 2　存储迁移原理图

Fig. 2　Schematic diagram of storage migration

CIMISS Oracle 数据库对应 70 个 LUN，总计需要 7000 个 IOPS，综合考虑缓存命中和硬盘提供的 IOPS、业务读写模型（读写比：9∶1）和读写惩罚等因素，本次为数据库规划 6 块 7.68 TB SSD（RIAD 10）和 40 块 2.4 TB 10K SAS 磁盘（RAID 6 8 +2 策略），预留 10% 空间用作自动分层迁移数据时用于中转空间使用。

VMware 虚拟化集群资源池占用主要存储空间，规划 16 块 7.68 TB SSD（RIAD 6）和 84 块 2.4 TB 10K SAS 磁盘，采用 RAID 6 8 +2 策略预留 8 块磁盘做应急性能和容量扩容使用；50 块 10 TB 7.2K NL SAS 323 TB 可用，预留 10% 空间用作自动分层迁移数据时用于中转空间使用。

3.3.2　数据迁移评估　数据迁移时间窗口 在目的存储接管源存储进行数据迁移时，不管是零中断迁移还是短暂中断迁移，都会对存储上层业务有一定的影响[6]。接管存储和数据迁移在双存储有大量的读写操作，IPOS 值是否能满足业务正常运行是重点要素。气象业务在台风、暴雨等天气过程时，业务繁忙，不适合进行数据迁移。选择数据迁移时间窗口原则是在业务空闲时存储负载量相对较小的时间段。

数据迁移存储 为保证业务的正常下发，在零中断迁移时需确保目的存储在源存储的空闲启动器至少有 2 个，且分配在不同的控制器上；不同型号与版本的目的存储接管的源存储的数量、LUN 的数量与路径数是有限制，充分考虑评估数据迁移的各种因素。

备份和配置 数据迁移之前一定要进行业务数据的备份操作，用于紧急情况下的数据回退，确保在执行一致性分裂时，选择业务较空闲阶段进行。在执行接管任务时，需要更改配置之后，目的存储才能接管存储，更改配置之前确认更改不会造成配置冲突。主机操作系统为 Linux 且使用的多路径软件为原生多路径（DMMultipath），在选择零中断迁移时，先完成主机多路径软件的配置。

3.3.3　风险评估和回退方案　业务数据迁移受多种因素影响，属于高危操作，必须做好气象业务风险评估和降低风险隐患的应急措施[7]（表 2）。

迁移风险。迁移前数据进行备份，以防止迁移失败可能导致的数据丢失；迁移前还需对源存储进行一次全面的检查，包括存储硬件以及存储的配置信息、性能等，确保满足迁移所需要的条件；迁移过程中，确保网络通畅，以防网络原因导致迁移失败。

回退方案。在执行数据迁移的过程中，遇到不可抗力因素或者其他影响数据迁移的因素，为了保证业务不中断，异常情况时可停止数据迁移操作，进行相应的回滚操作，确保业务的连续性以及业务数据的完成性。

3.3.4　数据迁移流程　数据迁移主要分为检查、创建任务、执行任务和验证 4 个流程，通过主机、业务、操作系统、配置和网络等状态检查，确保满足数据迁移的环境、配置、软件和网络要求；再通过创建任务、执行任务（业务接管、存储迁移）来完成数据迁移，任务结束后对主机、配置、系统和业务验证（图 3）。

验证是数据迁移的最后一步，也是判断数据是否迁移成功的重要准则[8]。查看主机链路状态，通过“lsdev －Cc hdiskx”查看所有磁盘列表和磁盘路径，状态为 Enable 则代表成功；登录 DeviceManager 查看告警中心，查看主机 IOPS、带宽、响应时间，无告警代表成功；验证系统性能报表查看性能报表，确认系统业务正常，且性能符合预期。

表 2　数据迁移风险评估

Tab. **2**　Data migration risk assessment form

序号	风险描述	发生可能性（高/中/低）	影响程度（高/中/低）	应对措施
1	LUN 数据丢失	低	高	严格按照计划执行并备份关键数据
2	eDevLUN 切换 IO 异常导致业务中断	低	高	根据错误码及其处理措施、收集日志分析原因处理
3	Zone 配置错误	低	中	备份光纤交换机 Zone 配置,以便可以快速恢复
4	Oracle 数据库异常及应对措施	低	高	数据库管理人员现场支撑变更,配合协助处理问题
5	SmartMigration 迁移失败	低	低	根据错误码及其处理措施、收集日志给研发分析原因处理
6	接管失败	低	中	决策回退方案

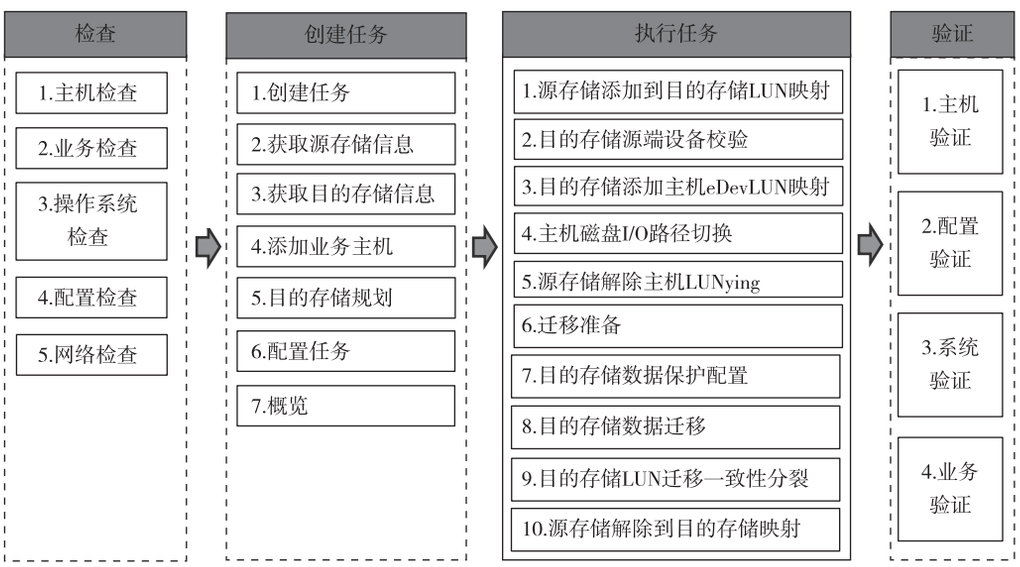

图 3　数据迁移流程图

Fig. 3　Data migration flow chart

3.3.5　数据迁移实现　本次迁移以业务空闲时存储负载量相对较小的时间段原则,通过 CIMISS 数据库、VMware 虚拟化资源池和 50 多个气象业务核心系统运维管理团队调研评估，选择 2021 年 11 月 15—19 日为数据迁移时间窗口,此时间段是在非汛期且无任何天气过程,广东省未发出暴雨、台风、寒冷、高温等预警信号,期间存储负载量是最高峰负载量的 1/8,也是全年存储负载量最低的时间段。

数据迁移准备。首先在 18810 V 5 新存储按规划完成逻辑卷 LUN 创建,进入存储管理系统分配 73 个 LUN 逻辑卷到主机,Hdisk4 – 76 是多路径软件纳管的磁盘,为数据存放提供优于原来的存储逻辑环境;其次通过对 2 套存储、气象业务系统主机逐一做健康检查,包括 CPU 占用、链路冗余、性能、网络、权限配置、主机多路径设备。

配置信息备份。存储多路径信息是数据迁移失败回滚的重要配置文件,备份配置信息有利于减少迁移失败恢复的时间。保存需要的虚拟磁盘属性信息,一一对应保存,用于升级后将磁盘属性重置回升级前的属性。其中包括虚拟磁盘选路算法、队列深度、预留策略等,通过执行命令 lsattr – El hdiskX 查询并保存虚拟磁盘的属性信息。

数据迁移。按原存储承载运行的 VMware 虚拟化集群资源池和 2 台服务器 CIMISS ORACLE 数据库类型的应用,分 2 次集中迁移,根据数据迁移流程逐一执行,动态观察、监控和验证,本次 430 TB 数据用时 82 h,在初始阶段新存储 0 负载的情况下最高速度可达 2.1 GB · s^{-1},随着数据的迁移,负载的增加,迁移速度也逐渐下降,其平均迁移速度为 1.5 GB · s^{-1}。

传统的数据迁移方法主要是借助相关工具在应用层上迁移。此种方法要求部署相对应的应用环境,适合单一气象业务系统以及存储数据少的应用场景。在应用层面迁移,经业务系统服务器的网卡、CPU、内存和磁盘处理,数据迁移速度降低到 350 ~ 750MB · s^{-1}。如果使用该方法,涉及的 CIMISS ORACLE 数据库,700 台虚拟机的超过 50 个气象关键核心系统都部署对应的应用环境,给气象业务系

统运维管理团队增加巨大的工作量。每个业务系统可迁移的时间窗口也不一样，迁移时间分散，预计迁移时间超过 1 个月，整套存储数据迁移时间长，造成数据迁移存在不确定性和增大迁移风险。

4　结语

气象业务具有其连续性、时效性等特点，支撑气象业务的底层存储的替换和数据平稳迁移要重点考虑业务影响和可操作性。同品牌、同家族系列的存储替换，在一定程度上降低了异构存储数据迁移存在的数据丢失、业务影响或中断的风险。基于气象业务零中断数据迁移，需要综合考虑多方面因素，包括所迁移的气象业务类型、数据量、气象业务系统连续性要求、可操作的有效时间窗口、气象业务系统的重要性程度、气象上下游业务量大小、源和目标存储是否是同构和数据迁移技术是否成熟等。当然，要成功实施一个基于气象业务零中断的数据迁移项目，不仅要选择成熟、合适、高效的数据迁移技术，更要通过严谨完整的规划和设计，迁移业务数据信息收集、迁移业务数据可行性分析、迁移业务数据风险评估、迁移业务数据方案验证、回退方案制定和迁移实施执行等环节缺一不可。

相比在应用层上做数据迁移的传统方法，基于存储底层块磁盘采用异构虚拟化迁移技术具有迁移速度快、时间短、应用层工作量少、业务无中断、可集中迁移的特点，适用气象业务系统类型多、数据量大以及时间要求短的场景。

参考文献

[1] 郑思轶，曾沁，胡胜，等. 广东省气象台 SWIFT 业务平台简介[J]. 广东气象，2018，40(2)：77－80.
[2] 李茂林. 负载均衡下的混合存储数据迁移方法研究[D]. 西安建筑科技大学，2020.
[3] 李刚，石艳，谭健，等. 贵州省智能化决策气象服务平台简介[J]. 中低纬山地气象，2021，45(1)：85－89.
[4] 刘宇. 数据迁移部署系统设计与优化研究[D]. 华南理工大学，2019.
[5] 张金标，李泽杰，乔文文，等. 基于分布式专有云的应用系统云化改造[J]. 广东气象，2018，40(6)：74－86.
[6] 杨俊萍，张广通. OpenMediaVault 存储方案在智能网格预报业务中的应用[J]. 中低纬山地气象，2018，42(5)：58－61.
[7] 刘敏，薛小宁，贺彩萍. 基于质量管理体系的基层综合气象业务管理分析[J]. 中低纬山地气象，2021，45(5)：122－124.
[8] 袁园，吴昆，顾今. 桌面云存储扩容项目中数据迁移方法研究[J]. 网络安全技术与应用，2018(2)，72－73.
[9] 陈旭辉，刘洋，高鹏，等. NVMe 在气象大数据分布式存储中的研究与应用[J]. 气象水文海洋仪器，2021，38(4)：12－15.
[10] 高建尽，张乾隆，武同元. 基于 Oracle 的分布式潮汐观测数据库系统的分析与设计[J]. 气象水文海洋仪器，2020，37(1)：40－44.

(上接第 110 页)

②在不同能见度等级的偏差特征分析中发现，能见度一般(2～10 km)等级和能见度较好(10～20 km)等级的一致性很好，能见度很好(20～50 km)等级的一致性较好，能见度很差(0～1 km)等级和能见度较差(1～2 km)等级的样本点较少，所以其分析欠缺一定的合理性。

③ 2 款能见度仪观测数值与相对湿度、雨量和气压相关性要比其他气象要素要高，与相对湿度和雨量均呈负相关，与气压呈正相关。

④无论晴天或雨天，相关系数日变化与气压日变化特征基本一致。在雨天，尤其是在高湿状态下，相关系数比低湿状态时要高。整体而言，2 款能见度仪在雨天时的日平均相关系数(0.93)比晴天(0.82)时要高。

参考文献

[1] 吴兑，廖碧婷，陈慧忠，等. 珠江三角洲地区的灰霾天气研究进展[J]. 气候与环境研究，2014，19(2)：248－264.
[2] 李蒙，刘文荣. 雾天多次散射对激光透射仪能见度测量的影响[J]. 激光技术，2020，44(4)：503－508.
[3] 甘桂华，张小荣. Belfort Model 6000 能见度仪工作原理与使用方法[J]. 广东科技，2012(23)：229－230.
[4] 张毅，刘小容，钟运新，等. 前向散射能见度仪的工作原理及维护维修[J]. 气象水文海洋仪器，2015，32(1)：118－120.
[5] 张顺，胡洋，周雨. 昌北机场能见度仪探测数据对比分析[J]. 气象水文海洋仪器，2020，37(4)：39－41.
[6] 李晓岚，权维俊，王东东，等. DNQ1 与 FD12 型能见度仪观测比对及影响因子研究[J]. 气象与环境学报，2021，37(3)：25－32.
[7] 佘元标，黄殷，郑贵生. 饶平站能见度仪器测量与人工观测对比分析[J]. 广东气象，2013，35(1)：77－80.
[8] 吴振强，李文斌，杨森槐，等. Belfort M6000 能见度传感器实测距离与人工目测距离的对比分析[J]. 科技创新导报，2011(32)：99－101.
[9] 马开玉，丁裕国，屠其璞，等. 气候统计原理与方法[M]. 北京：气象出版社，1993：29－30.
[10] 魏凤英. 现代气候统计诊断与预测技术(第 2 版)[M]. 北京：气象出版社，2007：18－19.
[11] PENG Y，PAN X，JIE T，et al. Hygroscopic growth of aerosol scattering coefficient：A comparative analysis between urban and suburban sites at winter in Beijing[J]. Particuology，2009，7(1)：52－60.
[12] CHEN J，ZHAO C S，MA N，et al. A parameterization of low visibilities for hazy days in the North China Plain[J]. Atmospheric Chemistry and Physics，2012，12(11)：4935－4950.
[13] XU W，KUANG Y，BIAN Y，et al. Current challenges in visibility improvement in southern China[J]. Environmental Science & Technology Letters，2020，7(6)：395－401.