气象私有云环境下存储架构设计与性能分析

王 彬,韩同欣,李 楠
(国家气象信息中心,北京 100081)

摘 要:为评估当前气象私有云平台中存储基础架构设计的合理性以及运行性能状况,并为将来扩充平台存储能力提供必需的架构级建设依据,从私有云所使用存储的存储协议、硬件性能等方面入手,在系统分析云计算环境下存储系统设计要点的基础上,给出了云环境下存储架构的设计思路,提出了云环境下结合不同业务数据读写特点的支持 SAN、NAS 多协议存储架构建设的思路,以构建统一存储资源池。结合实际搭建的气象私有云,分析其虚拟化环境下的存储应用性能。验证实验结果表明,当前环境中存储架构规划以及配置方式合理,存储各项性能运行指标良好,针对不同类型的应用均有适合的存储资源以供匹配,已建存储架构能够满足气象业务科研系统的存储需求。

关键词:气象私有云;存储性能设计;多协议存储架构;光纤通道存储;NAS 存储;分布式存储;存储资源池
中图分类号:TP39 **文献标识码**:A **文章编号**:1673-629X(2017)05-0020-05
doi:10.3969/j.issn.1673-629X.2017.05.005

Storage Architecture Design and Performance Analysis in Meteorological Private Cloud Environment

WANG Bin,HAN Tong-xin,LI Nan
(National Meteorological Information Center,Beijing 100081,China)

Abstract:For the evaluation of design reasonability and performance of the meteorological private cloud platform storage infrastructure, and to provide necessary architecture level basis for the future expansion of the storage capacity of the platform,on the basis of analyzing selection of storage protocols and hardware performance as well as discussion on the key issues of storage design in cloud computing environment,the thought for storage architecture design in cloud computing environment has been proposed. It is suggested that multi-protocol storage structure supporting SAN and NAS has been established in light of read-write features of various applications so as to build a standardized pool of storage resources. Combined with the actual meteorological private cloud,storage applications performance has been analyzed in virtualized environment. The verification experiment results show that the design and configuration of storage architecture are reasonable in current environment and various runtime performance indicators are good and the applications have been allocated suitable storage resources. The storage architecture of meteorological private cloud has been built to meet the storage needs of all kinds of meteorological operations and research systems.

Key words:meteorological private cloud;design of storage performance;multi-storage protocol;fibre channel storage;network attached storage;distributed storage;storage resource pool

0 引 言

云计算是近年来兴起并广受关注的一种资源提供、使用和计算模式:“云计算是由规模经济拖动,为互联网上的外部用户提供一组抽象的、虚拟化的、动态可扩展的、可管理的计算资源能力、存储能力、平台和服务的一种大规模分布式计算的聚合体”[1-2]。云计算环境中,任务作业分布在资源池中,各种应用系统能够根据需要实时获取计算能力、存储空间和各种基础软件服务,云计算平台可以按需对资源、平台和软件进行动态地部署、配置、重新配置以及取消等。云计算具有资源虚拟化、存储高效可靠、高可扩展性、集约管理、按需服务、“超瘦”客户端、使用方便等优点[3-5]。

相比于传统的 IT 系统建设与资源提供方式,云计算能够有效提升 IT 资源利用率,降低管理复杂度,加快 IT 响应速度。经过多年的发展,云计算已成为当前数据中心转型的最佳技术选择[6]。云计算和主机虚拟

收稿日期:2016-06-03 **修回日期**:2016-10-12 **网络出版时间**:2017-03-13
基金项目:科技部公益性行业(气象)科研专项项目(GYHY201306062)
作者简介:王 彬(1976-),男,正研级高级工程师,博士,CCF 会员(E200009018M),研究方向为云计算、高性能计算、气象信息化设计等。

化带来了计算和数据的大集中,为存储的性能、可靠性、可用性、可管理性等方面提出了挑战。为了实现这一目标,云计算平台中的存储系统设计显得至关重要。

结合业务实际需求,气象部门应用云计算技术建立了私有云计算环境,为用户提供计算资源、存储资源、网络接入与集中托管服务,满足用户对 IT 基础设施资源的需求。该系统利用云计算的按需、弹性、高可用服务等特点,提高了资源利用率,降低了建设及运行成本,实现了资源快速部署与动态分配,可随需求增长进行系统动态升级扩充。

根据建立气象私有云环境下存储的使用特性,通过对不同业务区域的存储进行实验,得出了关键性能指标,将分析数据与业界对于存储的性能所给出的参数指标进行了比对,并对其进行了深入分析,进而确定了气象私有云环境下存储的设计方案,并对架构及性能进行了评估。评估分析结果表明,已建成投入使用的气象私有云存储架构合理,能够很好地支撑气象业务科研系统的稳定运行。

1　云环境下存储设计关键点

随着云计算技术在 IT 架构变革中产生的效益逐步得到认可,企业或政府单位中的虚拟化架构承载的比例正在迅速增大。

气象部门的信息化建设已进入到信息技术与气象业务深度融合的阶段;在助力气象业务,有效提高“四个能力”的同时,以先进的设计理念、有效的组织形式和技术手段,尽可能提高工作效率和效益,是实现“又好又快”发展模式的主要途径[7]。国家气象信息中心基于云计算技术构建了气象私有云,为气象部门国家级业务单位提供计算、存储、网络接入等基础信息资源服务。“私有云”以 40 余台物理服务器以及 10 套磁盘阵列、NAS 存储为基础资源,对外提供 460 余台虚拟服务器,在其上运行了强天气预报、集合预报处理、雷达拼图、CIPAS、公服统计、气象业务内网、中国气象数据网、再分析评估等近 170 个应用系统。据初步测算,资源利用率提升 50%,业务部署上线时间从“月”缩短到“天”,故障恢复时间从小时级缩短到分钟级,CPU 利用率较此前提高了 6 倍以上,电力能耗节省和场地空间占用降低了 80% 以上。气象私有云改变了传统“一机一应用”的部署模式,提高了资源利用率,降低了建设及运行成本,实现了资源快速部署与动态分配,可随需求增长进行系统动态升级扩充[8]。

由于很好地满足了业务需要,气象私有云虚拟化资源规模快速增长。虚拟化服务器平台的不断扩大和承载业务的关键度不断提高、业务类型的复杂度不断增加,这些都对底层存储平台提出了新的要求。

1.1　业务实际需求

底层 IT 架构最终是为上层业务服务的。在以往的云平台建设中通常仅通过一种存储架构去适应所有类型应用产生的不同数据访问模式以及不同数据特征,这种方式在虚拟化架构规模较小、业务访问压力较低的场合下可以有效降低存储平台的设计难度。但随着云平台所承载业务的复杂度、关键度、访问量的增加,不同业务系统读写模式、数据类型的差异就逐渐被放大,因此需要为不同类型的业务选择最匹配其数据类型和数据访问特点的存储系统,从而优化存储平台的整体性能和综合成本。

云平台中的存储建设并不是一蹴而就的,可以通过以下手段对现有环境进行综合分析以逐步完善和优化:

(1)对具体业务虚机进行数据读写数据量和读写比例分析;

(2)利用主机虚拟化平台性能规划分析工具从虚拟化层得出数据吞吐性能;

(3)对现有存储前/后端口、缓存、磁盘组的历史性能进行统计分析。

1.2　存储设计要点

在对业务系统进行综合、全面的分析之后,就可以对存储平台进行有针对性的规划设计了。其中需要考虑的层面主要包括:如何进行存储协议的选择;如何针对不同业务数据类型解决性能问题;如何在多业务共享的情况下进行容量的统筹规划;如何在多台存储环境下实现存储平台的资源池化;如何针对不同类型存储实现存储平台的高可用保护。

2　云环境下存储协议的选择

选择合适的存储系统是虚拟化云计算环境整体架构设计的关键一环。云计算本身并未对存储架构做出明确限定,不同协议的存储在同一环境下的运行效果大相径庭。因此,在存储架构选择时,需要充分考虑存储在云平台中的用途以及所存储的数据类型。提供一个多协议的底层存储平台,根据不同数据类型的存储需求提供与之匹配的存储协议是当前主流的发展思路。图 1 是不同存储协议的应用比例[9]。

2.1　光纤通道存储

除去成本高之外,单从性能和可靠性的角度看,光纤通道(FC)无疑是当前最出色的存储协议。光纤通道当前的主流带宽为 8 Gbps、16 Gbps,其优势包括:

(1)具有更高的可用带宽、较低的延时和协议开销,通常情况下性能表现有保障;

(2)独立的光纤通道网络安全性更高,并有 Zoning 和 LUN masking 等访问控制机制;

(3)支持从存储启动系统(boot from SAN),服务器本地不再需要硬盘;

(4)基于 block 的块存储类型。

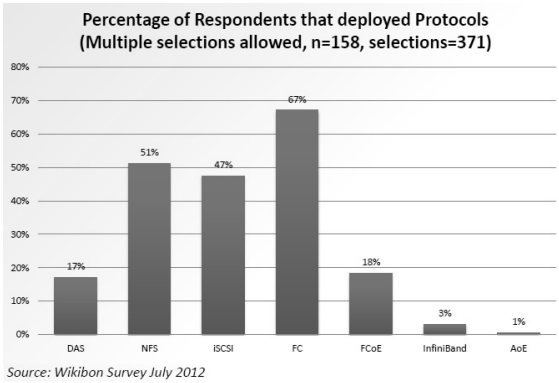

图 1 不同存储协议的应用比例(Wikibon Survey)

虚拟化架构下物理服务器上一般运行多个虚拟机,如业务系统对磁盘 I/O 有较高的要求。为了得到最佳的性能,首选使用基于光纤通道协议的存储系统。

2.2 NAS 存储

NAS 与 FC、iSCSI 之间最大的区别是协议类型不同。FC、iSCSI 使用数据块协议,数据以块为单位从前端虚拟机写入到后端的存储设备,主机端负责维护磁盘上的文件系统,在主机看来,存储设备与服务器本地磁盘并没有区别。NAS 则是另外一种情况,存储设备端负责维护磁盘文件系统,使用的是文件共享协议,服务器与存储设备之间只有简单的通讯。

一般说来,常规 NAS 设备的性能还达不到光纤通道存储的 I/O 并发能力和 I/O 低延迟,也无法支持主机虚拟化平台的全部存储特性,因此在大中型云计算场合下并不适合作为部署虚机操作系统的主存储。但是目前新兴的分布式 NAS 产品却可以有效发挥后端磁盘带宽,对于大容量的非结构化数据文件的高并发读写是最佳性能解决方案。因此在大型虚拟化、云计算场合内,完全可以采用一部分分布式 NAS 通过共享目录的方式来解决虚拟机中大文件的存储性能和存储成本问题。例如,部署在气象私有云上的气象再分析评估系统,虚机系统数据由 FC 盘阵提供存储空间,数据文件存储在分布式 NAS 中。

3 存储平台性能设计

云计算平台下的存储性能是整个云平台性能发挥的关键一环,如果设计考虑不周,将造成虚拟化主机的 I/O 大量积压,延时大大增加,严重时将导致上层应用连接超时中断或死锁。以下将从多个角度对气象私有云环境下的存储平台性能设计进行分析。

3.1 存储性能影响因素

云环境下对存储平台造成性能影响的不仅仅来自于存储控制器、磁盘类型等传统因素,还要考虑虚拟化主机环境与存储产品的集成度,只有通盘考虑才能保证存储平台的性能发挥。

(1)传统因素:与传统应用相一致,需要综合考虑存储产品的软硬件技术配置,例如存储控制器的处理器计算能力、存储缓存大小、磁盘转速及磁盘接口类型、IO 通道的带宽及数量、主机层面的多路径管理。

(2)加速盘(诸如闪存)的应用:闪存作为新兴的存储介质,由于其 I/O 性能是传统转轴硬盘的 30 ~ 50 倍以上,因此适合用于虚拟化环境中关键应用系统的性能提升。

(3)存储阵列和服务器虚拟化的集成:以往的服务器虚拟化产品在其自身的内核层承载了大量的 I/O 管理工作,随着虚拟化技术的发展,服务器虚拟化产品逐渐研发并开放了与存储产品兼容的软件接口协议,目前先进的存储产品均可与其对接并将大量的存储 I/O“卸载”到存储端,大大减少了服务器虚拟化层的压力,从而优化整个虚拟化环境的性能发挥。如气象私有云在虚拟机与存储设备之间采用了 VAAI、FusionSphere 等协议与主流存储系统进行对接,用户不需要关心后端存储的类型和能力。

3.2 存储 IO 能力计算

在真实的存储环境中,I/O 瓶颈往往来自于后端磁盘,因此衡量实际配置存储的 IOPS 应从磁盘系统规划入手,考虑到缓存性能优化作用的不确定性,最好在规划时不考虑或者尽可能少地考虑缓存的作用。根据存储系统规划理论,可套用如下公式(适用于磁盘阵列):

$$(IOPS * \%R + WP * IOPS * \%W)/\text{单盘 IOPS} = \text{所需磁盘数}$$

其中,IOPS 为存储系统 IOPS 需求;%R 为读操作百分比;%W 为写操作百分比;WP 为写惩罚因数,即 1 个写操作带来的磁盘 I/O 数,RAID1/0 和 RAID1 WP 为 2、RAID5 WP 为 4;单盘 IOPS 为单块磁盘的最大 IOPS 值(在存储设计过程中建议不采用磁盘厂商标称的单盘性能值,保守估计 10 K 磁盘约 125 IOPS;15 K 磁盘约为 175 IOPS;SSD 磁盘约为 3 000)[10];所需物理磁盘数为存储中应配置的最小磁盘数量。

根据业务系统使用特性,或者从以往业务主机抓取的 IO 分析数据,可以得出业务系统的大致读写比例,结合对 IOPS 的统计预期目标,即可计算出为实现此性能存储所需的磁盘数。

3.3 虚拟化层对性能的影响

随着虚拟化技术的发展,服务器虚拟化产品逐渐研发并开放了与存储产品兼容的软件接口协议,目前先进的存储产品均可与其对接并将大量的存储 I/O

“卸载”到存储端,大大减少了服务器虚拟化层的压力,从而优化整个虚拟化环境的性能发挥。

在虚拟服务器环境中,其存储硬件和 Hypervisor 管理程序的通讯非常复杂。为简化其通讯并提高效率,研发了 vStorage 阵列集成应用接口(VAAI)[11]。该应用接口为 Hypervisor 管理程序和存储设备规范了不同的职责,使其各自关注工作效能最大化,即 Hypervisor 致力于虚拟化相关的工作,而存储相关的工作则留给存储阵列。

通过 VAAI,存储阵列厂商可以直接将其存储硬件及应用程序和 vSphere 进行集成。VAAI 使得某些存储层的工作(诸如克隆等)可以在存储阵列上离线运行,较在主机端完成更为高效。主机端可以简单地将相关工作转到存储阵列上完成,而主机端只负责过程监控,而非使用主机端的资源来完成。存储阵列更擅长此类数据工作,可以较主机端更为快速地完成相关服务请求。

4　实验及其结果分析

对气象私有云环境中两块主要业务区域(A 区、B 区)的典型存储系统抓取了一天的完整性能数据,利用存储性能分析工具获取详细性能分析报表,作为优化存储平台性能的参考依据。

4.1　存储访问特性分析

不同类型的业务系统对存储的访问特性均有所不同,主要体现在 I/O 读写比例和 I/O 大小两个方面,存储性能优化需要充分考虑其访问特性。通过对气象私有云中的存储性能数据进行分析,A 区读写比例为 52% ∶48%,相对比较平均,I/O 大小以 4 K 为主;B 区两台存储读写比例分别为 60% ∶40% 和 33% ∶67%,其中 1 台存储设备具备了一定量的 128 K 以上大 I/O 的访问。具体如图 2 和图 3 所示。

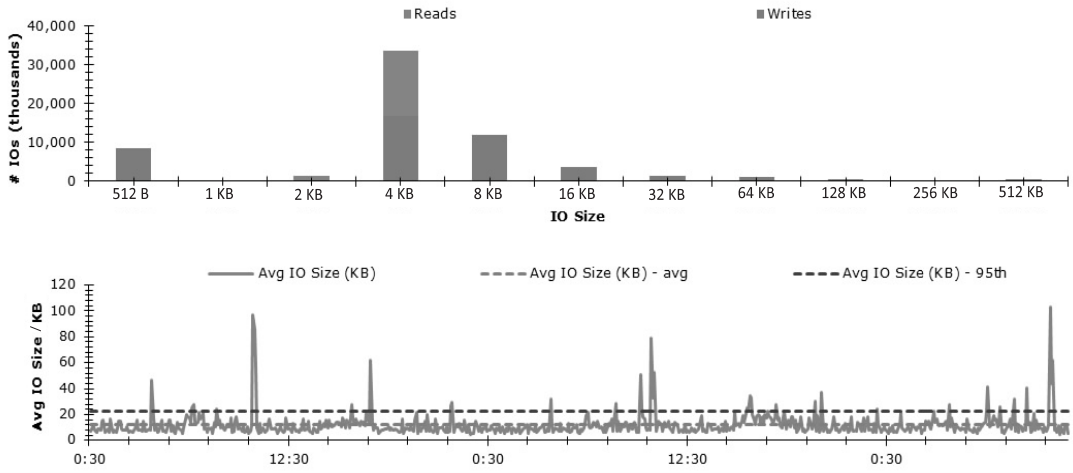

图 2　气象私有云 A 区存储 I/O 访问特性

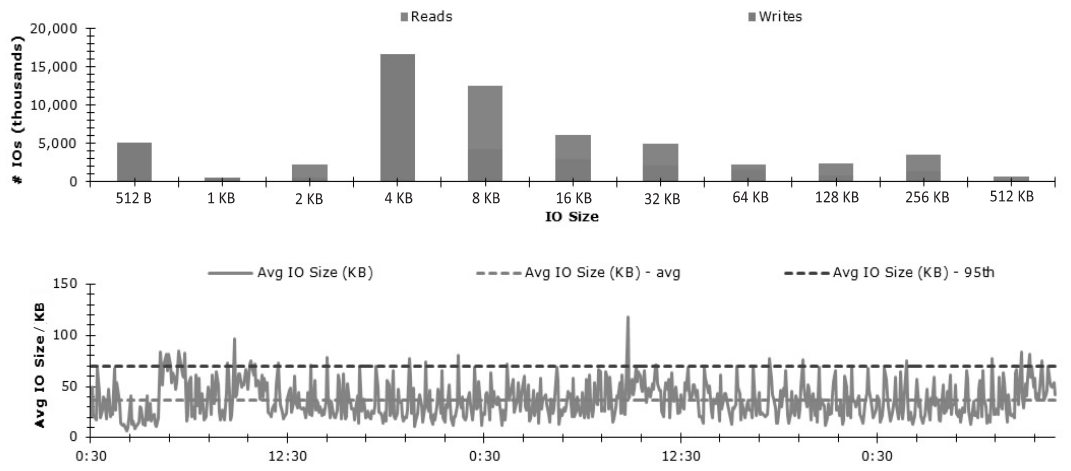

图 3　气象私有云 B 区存储 I/O 访问特性

对于以写入为主的存储访问方式,瓶颈往往出现在后端磁盘。因此在存储规划时应充分发挥后端磁盘的并发性能,数据卷尽可能打散分布在更多的物理硬盘上,提升数据落盘时的 I/O 响应能力。

气象私有云在设计时充分考虑了这一因素,设置存储后端磁盘以 Pool 的方式提供服务,数据卷打散在

Pool 内的所有硬盘上,突破了传统 RAID 组的性能局限。

对于大 I/O 业务量的场景,气象私有云存储可以通过虚机与业务数据分离部署的方式提升整体架构性能,即将虚机的系统文件部署在传统的光纤通道存储上,而将其业务数据目录指向高带宽、大容量的分布式 NAS,可以很好地提升大文件读写的性能,同时避免牵制云平台上虚机系统的性能。

4.2 存储负载性能分析

以下是几台存储的总体展现。从存储承担的 IOPS 上看,几台存储目前负载并不平均,A 区存储 IOPS 压力较小,B 区存储 IOPS 压力相对大。目前来看由于存储性能充足,95% 左右的 IO 响应时间均能控制在 5 ms 以内,属于非常理想的性能表现[12]。未来随着业务虚机数量的不断增长,存储 IOPS 压力不平均的问题就需要得到解决,否则部分存储有可能由于压力过大造成响应时间过长,影响了应用性能,而另一部分存储的性能却过于空闲无法充分发挥。

气象私有云 A 区和 B 区存储总体性能数据如图 4 和图 5 所示。

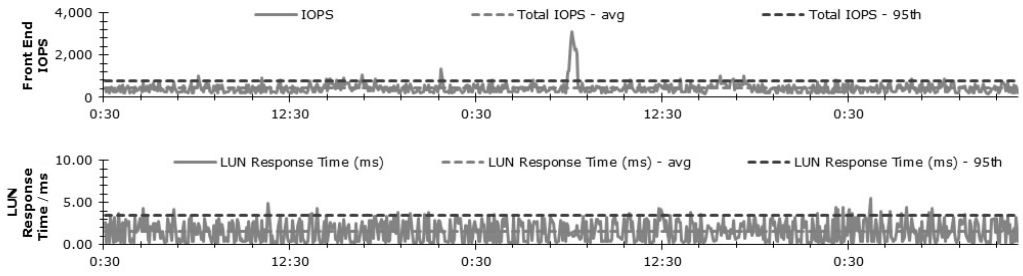

图 4 气象私有云 A 区存储总体性能数据

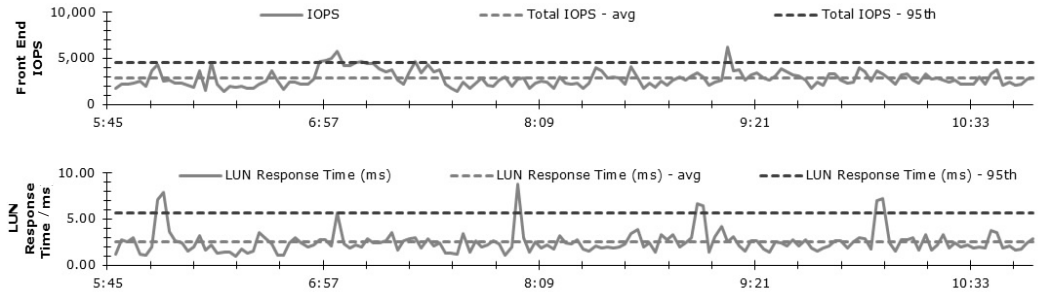

图 5 气象私有云 B 区存储总体性能数据

通过存储虚拟化技术可以实现存储资源的池化,很好地实现存储资源的动态灵活调整。基于存储虚拟化技术,通过进一步分析 LUN 性能热点,将压力较大的 LUN 在不影响业务应用的前提下在线迁移至相对空闲的存储,从而优化整体存储平台性能。

5 未来发展思路

气象私有云的建设是一个循序渐进的过程。在未来发展过程中,为了更好地满足各类气象业务可研系统的多样化存储服务需求[13-15],在横向扩展计算资源池的同时,将从以下两方面进行深化建设:

(1)随着云平台上业务系统的不断增加、数据量不断增大,需要根据业务系统 I/O 性能需求以及数据格式的不同,针对其存储特点构建多元化的存储平台。采用高性能的全闪存磁盘阵列来部署对 I/O 性能要求最为苛刻的应用数据,采用大容量的分布式 NAS 存储产品来部署气象业务系统中诸如模式数据、卫星图片等需要共享的类文件数据等,使应用得到最佳的存储服务级别。

(2)随着存储平台向多元化、虚拟化方向发展,对于庞大的低层存储架构需增加更为强大的统一存储管理,能够彻底跨越异构存储产品带来的管理鸿沟,为上层应用系统提供强大、灵活、优化的存储供给服务。将异构的 SAN、NAS、HDFS 分布式文件系统等存储对象抽象为统一的虚拟化存储池,通过统一的管理界面提供数据块、文件、分布式文件系统等多种存储协议,屏蔽底层存储设备管理的差异性,实现对底层各存储及存域网络设备的集中管理、维护、变更操作,并通过自动化技术进一步加速存储容量分配和配置变更效率。

6 结束语

围绕气象私有云平台存储设计与业务应用,分别从性能影响因素、IO 能力计算以及虚拟化层影响等方面进行了分析设计,同时评估了存储实际运行状况,从存储平台访问、IO 负载等角度进行了实验及分析。结

(下转第 29 页)

5　结束语

针对交通仿真系统中可能出现的死锁现象,通过进行多方位分析,提出了建设性的解决思路,并优化设计算法,在基于系统结构可嵌入的原则下,实现了一种新的死锁处理算法,使得交通仿真系统能够避免大多数的死锁情况发生。

交通的实际情况总是千变万化。对于交通仿真系统来说,所提出的算法可以根据出现的新情况新问题研究相应的对策,最大程度上保证交通管理在不需要用户干预的情况下,自动、高效地完成交通的自动控制任务。

参考文献:

[1] Roozemond D A. Using intelligent agents for pro-active, real-time urban intersection control[J]. European Journal of Operational Research, 2001, 131(2): 293-301.

[2] Chaib-Draa B, Moulin B, Mandiau R, et al. Trends in distributed artificial intelligence[J]. Artificial Intelligence Review, 1992, 6(1): 35-66.

[3] 张飞舟,曹学军,孙　敏. 基于多智能体的城市交通集成控制系统设计[J]. 北京大学学报:自然科学版,2008,44(2):289-292.

[4] 郭建钢,伍雄斌. 多智能体技术在交通系统协调控制中的应用[J]. 华东交通大学学报,2005,22(6):38-41.

[5] 吴继伟,杨定鹏,萧蕴诗. 多智能体协作方法及其应用研究[J]. 控制与决策,2004,19(2):216-218.

[6] Almejalli K, Dahal K, Hossain A. An intelligent multi-agent approach for road traffic management systems[C]//Control applications, intelligent control. [s. l.]: IEEE, 2009: 825-830.

[7] 何　涛,白振兴. 多智能体系统设计的关键技术研究[J]. 现代电子技术,2006,29(14):31-34.

[8] Hirankitti V, Krohkae J, Hogger C. A multi-agent approach for intelligent traffic-light control[J]. World Congress on Engineering, 2007, 79(3): 116-121.

[9] 王龙飞,陈　红,李　扬,等. 多智能体在城市交通系统中应用现状综述[J]. 计算机系统应用,2010,19(1):198-203.

[10] 史　乐,李　辉,原江波. 基于消息通信的多智能体系统的应用[J]. 计算机应用,2008,28(2):531-534.

[11] 贺　雷,刘正熙,毋攀良. 基于通用触发器系统的地面交通仿真[J]. 计算机应用,2007,27(11):2623-2625.

[12] 吴　越,周学农. 智能协作技术在交通管理中的应用[J]. 系统工程,2001,19(1):52-55.

[13] 欧海涛,张卫东,张文渊,等. 基于多智能体技术的城市智能交通控制系统[J]. 电子学报,2000,28(12):52-55.

[14] 李振龙,赵晓华. 基于 Agent 的区域交通信号协调控制[J]. 武汉理工大学学报:交通科学与工程版,2008,32(1):130-133.

(上接第 24 页)

果表明,气象私有云平台已建存储架构设计合理,不同类型存储经过优化配置后优势互补,能够满足各类气象业务科研系统的存储需求,实际运行监控指标良好。未来还将跟踪技术发展趋势,结合业务系统发展需要,在多元化存储和统一存储管理平台等方面做进一步研究,并应用到实际工作中。

参考文献:

[1] Foster I, Zhao Y, Raicu I, et al. Cloud computing and grid computing 360-degree compared[C]//Grid computing environments workshop. [s. l.]: IEEE, 2008: 1-10.

[2] 陈　康,郑纬民. 云计算:系统实例与研究现状[J]. 软件学报,2009,20(5):1337-1348.

[3] 刘正伟,文中领,张海涛. 云计算和云数据管理技术[J]. 计算机研究与发展,2012,49(S1):26-31.

[4] 陈　全,邓倩妮. 云计算及其关键技术[J]. 计算机应用,2009,29(9):2562-2567.

[5] 王意洁,孙伟东,周　松,等. 云计算环境下的分布存储关键技术[J]. 软件学报,2012,23(4):962-986.

[6] 沈文海. 气象业务信息系统未来基础架构探讨-"云计算"和"大数据"在气象信息化中的作用[J]. 气象科技进展,2015(2):64-66.

[7] 沈文海. 从云计算看气象部门未来的信息化趋势[J]. 气象科技进展,2012(2):49-56.

[8] "气象私有云":我们身边的云计算[EB/OL]. 2014-08-13. http://www.cma.gov.cn/kppd/kppdkjzg/201408/t20140813_257125.html.

[9] Wikibon Survey[EB/OL]. 2012-08-23. http://wikibon.org/wiki/v/VMware_vSphere_5_Users_Move_Beyond_the_Storage_Protocol_Debate.

[10] Getting the hang of IOPS v1.3[EB/OL]. 2012-01-28. http://www.symantec.com/connect/articles/getting-hang-iops-v13.

[11] VMwarev Sphere Storage APIs-Array Integration (VAAI)[EB/OL]. 2013-03-17. http://www.vmware.com/resources/techresources/10337.

[12] What'san acceptable I/O latency? [EB/OL]. 2010-09-19. http://kaminario.com/company/blog/whats-an-acceptable-io-latency/.

[13] 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.

[14] 吴焕萍,张永强,孙家民,等. 气候信息交互显示与分析平台(CIPAS)设计与实现[J]. 应用气象学报,2013,24(5):631-640.

[15] 王　彬,周　斌,魏　敏. 气象计算网格模式预报系统的建立与优化[J]. 计算机应用研究,2010,27(11):4182-4184.