

虚拟化绿色智慧交通数据中心关键技术研究

贾朝龙¹⁻³, 郝晓楠^{1,3}, 王雪纯^{1,3}, 曾友渝^{1,3}

(1.重庆邮电大学软件工程学院, 重庆 400065;

2.田纳西大学诺克斯维尔校区, 田纳西州 诺克斯维尔 37996;

3.重庆市软件质量保证与测评工程技术研究中心, 重庆 400065)

摘要: 采用基于VMware vSphere的数据中心虚拟化体系方案设计和SDN的高并发多应用SLA的资源管理方法, 对数据中心能耗建模及优化, 简化基础设施和资源管理方式, 减少资源开销, 整合资源, 节约建设成本, 实现智能、灵活、集约的业务调度, 实现资源按需分配, 提高智慧交通数据中心能源效率、资源利用率和设备能效比, 达到绿色节能。

关键词: 智慧交通; 虚拟化; 数据中心; VMware vSphere; 大数据

中图分类号: TP391.1 **文献标识码:** A

Research on Key Technologies of the Green Intelligent Transportation Data Center

JIA Chaolong¹⁻³, HAO Xiaonan^{1,3}, WANG Xuechun^{1,3}, ZENG Youyu^{1,3}

(1.School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2.The University of Tennessee at Knoxville, Knoxville 37996, USA;

3.Testing and Assessment, Chongqing Engineering Research Center of Software Quality Assurance, Chongqing 400065, China)

Abstract: This paper adopts the data management scheme of data center virtualization system based on VMware vSphere and SDN's high concurrency & multi-application SLA resource management method. It realizes the modeling and optimization of energy consumption in data center, simplifies the infrastructure and resource management, reduces resource cost, integrates resources, saves construction costs to achieve intelligent, flexible and intensive business scheduling, to achieve the allocation of resources on demand, to improve the energy efficiency of intelligent transportation data center and the efficiency of resource utilization as well as equipment energy efficiency to reach the goal of green energy.

Keywords: intelligent transportation; virtualization; data center; VMware vSphere; big data

1 引言(Introduction)

在智慧交通大数据中心方面, 由于交通运输行业是数据密集型行业, 大量的交通数据必然要求高效的管理和研究分析手段。智慧交通大数据中心建设已经是迫切需要, 并且智慧交通大数据中心构建必须把我当前数据中心技术发展的趋势。

在智慧交通大数据中心构建方面, 国内外许多学者进行了大量的研究工作。Du^[1]提出了一种新颖的城市交通数据即服务(CTDaaS), 融合来自分布式提供商的数据, 构建了一个因特网的流量数据服务(IoTDS)模型, 以识别数据资源之间的关联和关系, 并在透明计算范式和面向服务的架构下开发了CTDaaS代理, 根据不同的计算模型融合来自各种数据源的知识, 以及响应差异化的数据质量(QoD)。Jeon^[2]提出一种新的统计建模方法, 该方法根据每个链路的各种分析找到最佳的历史数据集, 并按每周七天提供更精确的交通流预测, 采用

基于长点分析、相关分析和蒙特卡罗模拟的三步过滤算法, 并通过使用诸如均方误差(MSE)和Akaike信息准则的决定因素来确定最佳历史数据范围, 构建了一个大数据处理框架来处理整体预测过程和计算大量的交通数据。Xiong^[3]提出深圳交通系统(SZTS, 一个新的大数据Hadoop基准套件), 包括现实生活交通分析应用程序与深圳的现实生活输入数据集, 专注特定的和真实的应用领域, 在微架构级别, 操作系统级别和作业级别执行跨层工作负载表征, 显示SZTS与现有Hadoop基准, 以及通用多核PARSEC基准相比的独特特性。同时还研究工作负载行为相对于输入数据大小的敏感性, 并提出了一种用于识别代表性输入数据集的方法。Huang^[4]提出了一种基于云计算技术的安全并行地图匹配系统, 通过在MapReduce范式中重构来适应云计算环境的串行跳跃映射匹配算法, 在混合云上的隐私感知地图匹配模型, 以实现敏感的GPS数据保

护，并在hadoop平台上实施1000亿记录的大型车辆跟踪数据集测，证明是高效率的大规模车辆跟踪数据处理。

2 数据中心与数据中心虚拟化(Data center and data center virtualization)

2.1 传统数据中心面临的机遇与挑战

数据中心起源于20世纪60年代以数据存储和简单计算阶段的计算中心，20世纪80年代逐步发展为以数据处理及业务应用阶段的信息中心，随后在21世纪初出现了以服务为主导的IT服务中心。随着云计算和虚拟化的快速发展，通过虚拟化技术搭建下一代数据中心被越来越多的采用，虚拟化技术为数据中心在架构、系统、业务模式和管理上提供一种全新的解决方案。

高效的数据查询检索性能是现代数据中心的一个重要指标，传统数据查询技术特征如表1所示。随着技术的发展，传统数据中心采用关系数据库查询、串行查询，以及后来出现的倒序索引、并行索引已经不能满足大数据时代发展要求，需要更先进的检索查询技术。

表1 传统数据查询技术特征

Tab.1 Traditional data query technology features

查询手段	技术特征	效率
传统的关系数据库查询	将元数据存入关系数据库，面对用户的多条件属性查询，逐条检索	效率较低
串行查询	各数据单位各自返回查询结果，每个单位内部查询仍然是串行查询	效率较低
倒序索引	多个属性联合查询	提高查询效率
并行索引	索引阶段和查询阶段都是并行处理	提高查询效率

2.2 虚拟化绿色数据中心

针对传统数据中心利用率低、缺乏灵活性、业务连续性差、可持续发展能力不足、运营能耗成本高、运维管理水平不高、自动化程度低、绩效评估难的不足，虚拟化技术释放了数据中心的潜力，为数据中心带来了极大的运营弹性。未来数据中心的面貌就是将底层的硬件，包括服务器、储存与网络设备全面虚拟化，建立起一个共享的按需而选的运作环境，上层的数据可以根据业务形态的不同需求，搭配出各种互相隔离的应用，形成一个服务导向的IT架构。虚拟化技术将更先进的自动化和系统管理引进到数据中心，为数据中心提供了一种共享加隔离的安全、可信的云计算基础架构，成为叠加了云计算功能架构的弹性数据中心。虚拟化数据中心的技術优势主要体现在提升IT设备利用率、简化管理、快速部署业务、更敏捷地支撑业务发展、绿色节能、实现高效的IT治理五方面。目前虚拟化数据中心正朝着分布式建设模式，数据中心管理工具(DCIM)的重构和发展为核心的控制点，软件定义数据中心，数据向集中化管理方面发展且可靠性成为未来数据中心的基础能力，更强调能源的有效利用与成本收益的优化几个主要方面发展。

VMware vsphere Hypervisor是一款免费的裸机

hypervisor，能够虚拟化服务器，以便将应用整合到更少的硬件上。在内置管理工具方面，可在几分钟内轻松创建和调配虚拟机。在存储使用效率方面，可以超出物理存储的实际容量过量分配存储资源。在高级内存管理方面，可以超额分配内存资源并执行页面共享和压缩，从而优化内存资源的性能。另外，它还可以经过强化的驱动程序可实现高可靠性，通过与独立硬件供应商合作，确保实现vSphere Hypervisor的最佳性能。

IDC报告表明，通过虚拟化和现有安装服务器重新部署与管理程序相结合，全球每年避免新增大量的服务器。由于使用VMware而避免新增服务器，虚拟化产品从2003年的总计107000个增长到2016年的1620万个。全球服务器虚拟化从2003年的107000增加到2016年的5350万。由于使用了VMware虚拟化产品而避免了排放量。超融合基础设施和软件定义网络进一步减少基础架构设备(存储阵列和网络)数量的新增。由于使用VMware虚拟化产品，每年避免增加的基础设施设备，从2014年总计595000台增长到2016年的107万台。

3 基于VMware vSphere虚拟化智慧交通数据中心 (Virtualized smart transportation data center based on VMware vSphere)

3.1 问题与现状

交通运输行业的业务系统建设通常以满足单一业务部门需求为主，具有业务覆盖单一、系统重复建设，以及系统之间相互割裂的特点，形成一个个信息“孤岛”和“烟囱”。交通运输行业中IT资源浪费严重与发展不平衡导致目前业务系统信息化发展水平不能满足交通行业高速发展的需要。同时，不同部门信息化存在巨大差距，部分交通信息化资源的配置和使用缺乏有效的监管，加之业务平台具有管理复杂、业务上线时间长的特征，导致业务连续性无法保障，影响了交通信息化的发展。随着数据量的增多，计算、存储密度增大，数据存储平台对安全性、扩展性也提出更高的要求，急需一套安全可靠的数据存储平台，保障数据的安全。

3.2 交通数据中心虚拟化

对于交通部门希望实现高效绿色虚拟化交通数据中心来说，VMware vsphere Hypervisor因其具有高度可配置性，对于要求完全虚拟化或选择混合方法的交通部门来说，它可以成为一个有吸引力的选择。以VMWare vSphere 5.5 Hypervisor来说，其性能参数如表2所示。

表2 VMWare vSphere 5.5 Hypervisor性能参数

Tab.2 VMWare vSphere 5.5 Hypervisor performance parameters

主机		虚拟机		集群			
逻辑处理器	物理内存	每个主机的虚拟CPU	每个虚拟机的虚拟CPU	每个虚拟机的内存	每个主机活动虚拟机	最大结点数	最多虚拟机
320	4TB	4096	64	1TB	512	32	4000

交通数据中心虚拟化将计算、存储、网络、大数据分析、智能分析、灾备、安全等系统,以及视频设备、网络设备、服务器、中间件、数据库、业务应用和机房动力环境等分割管理的信息孤岛进行有效的整合和关联,实现数据中心全网资源的全面监控与集中统一管理。根据计算需求,采用服务器虚拟化应用,对于视频、图像数据采用云存储模式,对于数据库、虚拟机镜像文件等数据采用FC-SAN共享存储。

对于交通大数据分析和智能分析需求,采用大数据平台针对非结构化、半结构化,以及结构化的过车视频、数据进行建模分析和交通数据采集,支持全文快速检索、以图搜图、智能研判、OD分析和统计分析等应用。对于灾备需求,可基于存储复制、双机热备、异地容灾等方式,进行两地三中心、同城双中心或异地双中心的容灾建设,同时可选用NBU备份或CommVault备份方案实现数据备份。

交通数据中心虚拟化从虚拟化方案的拓扑结构设计,虚拟架构主体Vmware ESX Server配置、SAN集中存储实现虚拟服务器的文件共享、虚拟架构环境的集中管理、自动化及优化运行,以及虚拟架构环境的整合备份四方面进行,其架构如图1所示。

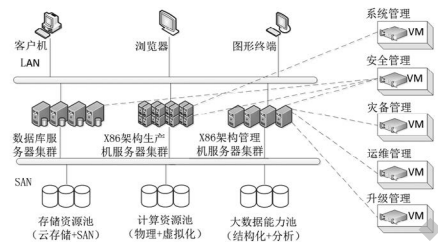


图1 虚拟化数据中心虚拟化架构

Fig.1 Virtualized data center virtualization architecture

4 虚拟化智慧交通数据中心关键技术(Key technologies of virtualized smart transportation data center)

4.1 动态交通监控数据存储及数据容错

在交通监控领域,监控数据分析的效率决定价值,因此需要更低的延迟、更准确的分析。随着数据量的增加,即使对TB级别的数据进行对视频内容的数据分析和检索,采用串行计算的模式都可能需要花费数小时的计算,已远远不能胜任时效性的需求。大数据架构下的存储系统还需要考虑后续的计算模式的匹配。通过分布式资源管理框架,结合存储区域网络(SAN)和软件定义存储(SDS),考虑动态交通监控数据的元组与属性特点,结合不同应用环境分别进行建模,并基于智慧交通业务中各子任务映射关系特点,分析各业务的容错约束条件及参数变化,考虑副本技术与元数据服务,基于集群的中间数据容错机制,将容错设计到架构之中,以可扩展的分布式系统的方式实现。

4.2 高并发多应用SLA的资源管理

基于高并发多应用SLA(Service Level Agreement)的兼容性和基于进程粒度的共享平台资源管理^[5],研究根据应用资源需求进行应用到物理机的应用放置/迁移和动态分配物理

机资源,降低SLA受到影响或资源浪费概率的发生的同结点应用资源的动态调整;基于虚拟机粒度的虚拟化平台资源管理,研究包括虚拟机的放置/迁移和同结点虚拟机资源的动态调整。由于虚拟机中应用负载具有随时变动性,借助虚拟化技术提供的细粒度资源分配机制支持,为虚拟机动态增加或减少资源。

4.3 数据中心能耗建模、优化和管理

对虚拟化数据中心的能耗管理从能耗监控与测量、能耗分析与建模、能耗管理实现机制,以及能耗管理优化算法四个方面进行研究。利用客户虚拟机的能耗管理策略实现虚拟机的能耗管理,以及利用硬件能耗管理机制实现虚拟机的能耗管理两种传统解决方法的不足。结合软件和硬件层次上能耗调节技术的方法来进行虚拟化云计算平台的能耗管理,包括虚拟化云平台的能耗分析与建模、服务器整合建模,以及在线迁移建模。

4.4 高效查询架构

高效查询架构采用Lucene和Hadoop, Lucene实现倒序索引,提供全文检索,并行索引阶段和查询阶段用Hadoop的Map和Reduce过程实现,显著提高查询效率。数据及时交由实时计算框架进行处理,同时定期同步至离线计算框架;实时计算框架处理接收到的实时数据,并将处理结果输出到数据查询框架或者离线计算框架;离线计算框架则定期对数据进行处理,并将处理结果输出至数据查询框架;数据查询框架对海量数据进行统计、汇总和分析,发现重复发生概率较高的模式,并通过应用云面向用户提供服务。

5 虚拟化绿色智慧交通数据中心设计实现(Virtualization of green smart transportation data center design and implementation)

5.1 虚拟化绿色智慧交通数据中心技术路线

虚拟化绿色智慧交通数据中心分三层结构,底层是前端感知层,负责将各类检测数据收集汇总,形成交通大数据;弹性计算层通过对数据预处理、分析及优化通过虚拟化存储、网络和计算平台实现安全、可用和自动化;业务应用层负责将各类检测数据收集汇总,形成交通大数据;弹性计算层通过对数据预处理、分析及优化通过虚拟化存储、网络和计算平台实现安全、可用和自动化;业务应用层负责将各类检测数据收集汇总,形成交通大数据;弹性计算层通过对数据预处理、分析及优化通过虚拟化存储、网络和计算平台实现安全、可用和自动化。

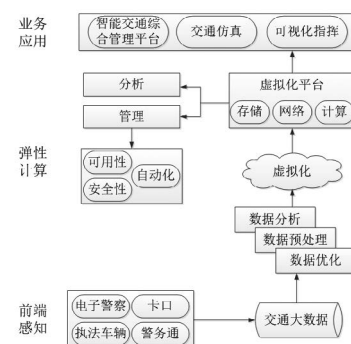


图2 虚拟化智慧交通数据中心技术路线

Fig.2 Virtualized smart traffic data center technology route

5.2 虚拟化绿色智慧交通数据中心方案

虚拟化绿色智慧交通数据中心组成包括五部分,分别是

内网数据、内网虚拟化平台、内网应用、外网虚拟化平台、应用需求。方案如图3所示。

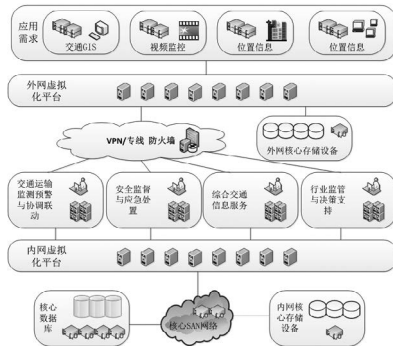


图3 虚拟化智慧交通数据中心方案

Fig.3 Virtualized smart transportation data center solution

6 结论(Conclusion)

通过交通数据中心虚拟化，实现现有物理资源的集中管理和统一调度，提高了交通部门设备利用率和能源利用的有效性，提升运营效率，以及降低投资开支。选用VMware虚拟化产品提高了交通数据中心系统基础构架的自动化程度和灵活性，满足了不断扩大的业务需求。同时，虚拟化数据中心不间断运营和集成化管理工具可以简化运营，加快解决问题的速度，并有助于向新技术转型。

(上接第25页)

如图3所示，直升机制造关键设备健康的大数据可视化分析的处理流程如下：

步骤1：收集数据。将直升机制造关键设备的原始数据序列化到hadoop中的HDFS中，并持久化到HBase非结构化数据库。

步骤2：根据业务需求在算法工具箱中选用相应的算法对数据进行挖掘。

步骤3：将数据挖掘算法输出的结果集映射到可视化模块中成为图形信息，可视化引擎将结果集与场景集成，将可视化效果图反馈给用户。

6 结论(Conclusion)

大数据是实现工业4.0的技术基础，在直升机制造中的应用中，通过建立大数据存储、数据分析等模型，能够对直升机零部件生产加工全工序质量信息进行自动化采集、集成管理与质量预警。充分发挥了大数据技术在数据处理能力中的优势，有效完善了直升机制造和装配中的建模，以及数据的处理，从而确保了直升机制造的加工效率和产品质量。

参考文献(References)

- [1] Baesens,B.,Bapna,R.,Marsden,et al.Transformational issues of big data and analytics in networked business[J].MIS Quarterly,2016,40(4):807-818.
- [2] Robert Thorstad,Phillip Wolff.A big data analysis of the relationship between future thinking and decision-making[J].PNAS February 20,2018,115(8):E1740-E1748.
- [3] Chaowei Yang,Qunying Huang,Zhenlong Li,et al.Big Data and cloud computing: innovation opportunities and challenges[J].International Journal of Digital Earth,2016,10(1):13-53.
- [4] Alexandros Labrinidis,H.V.Jagadish.Challenges and opportunities with big data[J].Proceedings of the VLDB Endowment

参考文献(References)

- [1] Du B,Huang R,Chen X,et al.Active CTDaaS:A Data Service Framework Based on Transparent IoD in City Traffic[J].IEEE Transactions on Computers,2016,65:1.
- [2] Jeon S,Hong B.Monte Carlo simulation-based traffic speed forecasting using historical big data[J].Future Generation Computer Systems,2016,65:182-195.
- [3] Xiong W,Yu Z,Eeckhout L,et al.ShenZhen transportation system(SZTS):a novel big data benchmark suite[J].Journal of Supercomputing,2016,72(11):1-28.
- [4] Huang J,Qie J,Liu C,et al.Cloud computing-based map-matching for transportation data center[J].Electronic Commerce Research & Applications,2015,14(6):431-443.
- [5] 张伟,宋莹,阮利,等.面向Internet数据中心的资源管理,软件学报,2012,23(2):179-199.

作者简介:

贾朝龙(1979-),男,博士,副教授.研究领域:大数据及可视化.

郝晓楠(1995-),男,本科生.研究领域:软件开发.

王雪纯(1996-),女,本科生.研究领域:软件开发.

曾友渝(1996-),女,本科生.研究领域:人工智能.

nt,2012,152(10):2032-2033.

- [5] Chang F,Dean J,Ghemawat S,et al.Big Table:A distributed storage system for structured data[J].ACM Transactions on Computer Systems,2008,26(2):1-26.
- [6] Ghemawat S,Gob IOFF H,Leung P T.The Google file system[C].Proceedings of the 19th ACM Symposium on Operating Systems Principles.New York:ACM Press,2003,29(8):29-43.
- [7] Dean J,Ghemawat S.Map/Reduce:Simplified data processing on large clusters[C].OSD Ip04:Proceedings of the 6th Symposium on Operating System Design and Implementation.New York:ACM Press,2004,19(6):137-150.
- [8] 黄阳华.德国“工业4.0”计划及其对我国产业创新的启示[J].经济社会体制比较,2015(2):1-10.
- [9] 刘智慧,张泉灵.大数据技术研究综述[J].浙江大学学报:工学版,2014(06):957-972.
- [10] Zhang Guanglan,Sun Jing,Chitkushev L,et al.Big data analytics in immunology:a knowledge-based approach[J].Biomed Research International,2014(1):9-18.
- [11] 李德仁,李德毅.论空间数据挖掘和知识发现的理论与方法[J].武汉大学学报:信息科学版,2002(03):221-233.
- [12] Wu Yingbin,Guo Fuliang.Visualization technology based on the treatment of complicated data[J].Computer & Digital Engineering,2015,43(11):1985-1989.

作者简介:

张贝贝(1991-),女,硕士,工程师.研究领域:大数据,人工智能及云计算.