

面向数据的云计算研究及应用综述

葛文双, 郑和芳, 刘天龙, 马 昭, 张瑞权, 吴成晟

(中国电子信息产业集团有限公司第六研究所, 北京 100083)

摘要: 云计算作为新一代信息技术, 其集成了许多先进的技术实践, 并充分利用多样化数据和强大计算能力, 进而实现用户业务系统的自适应。从数据角度出发, 针对数据密集型数据中心进行多层次对比研究, 综述了该场景下云计算的研究和应用现状。除了概述虚拟化技术、分布式计算模型、海量数据存储与管理技术、云管理平台等关键技术外, 提出了运用数据来源分析云计算应用的方法, 将其分为商业云、工业云、特种行业云, 并重点介绍了相应架构和应用特点。提出了下一步的研究方向, 为云计算相关研究提供一些参考。

关键词: 云计算; 数据中心; 虚拟化; 工业云; 特种行业云

中图分类号: TP301

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.200306

中文引用格式: 葛文双, 郑和芳, 刘天龙, 等. 面向数据的云计算研究及应用综述[J]. 电子技术应用, 2020, 46(8): 46-53.

英文引用格式: Ge Wenshuang, Zheng Hefang, Liu Tianlong, et al. Overview of data-oriented cloud computing research and application[J]. Application of Electronic Technique, 2020, 46(8): 46-53.

Overview of data-oriented cloud computing research and application

Ge Wenshuang, Zheng Hefang, Liu Tianlong, Ma Zhao, Zhang Ruiquan, Wu Chengsheng

(The 6th Research Institute of China Electronics Corporation, Beijing 100083, China)

Abstract: As a new generation of information technology, cloud computing based on diverse data and powerful computing capabilities. It integrates some advanced technical practices, and some self-adaption business systems are achieved. From the perspective of data, this paper developed multi-level comparative research on data centers, and summarized the current research and application of cloud computing. Besides an overview of key technologies such as virtualized technology, distributed computing models, massive data storage and management technologies, this paper proposed a method for analyzing cloud computing applications based on data sources, which are divided into internet clouds, industrial clouds, special field clouds. And this paper focused on the corresponding architecture and application features. Finally, the next research directions are provided and this could be helpful for cloud computing related research.

Key words: cloud computing; data center; virtualization; industry clouds; special field clouds

0 引言

随着信息技术的不断升级, 业务种类的增多, 各领域需要处理的数据呈现出数据量大(Volume)、数据生产速度快(Velocity)、多样性(Variety)及真实性(Veracity)“4V”特点。一般地, 具备“4V”特征的数据集都可认为是大数据。而随着大数据时代的到来, 面对海量的数据信息时, 传统的数据分析技术无法完全挖掘其中蕴含的信息^[1], 基于此, 云计算概念应运而生。另外, 研究人员一直在探索的人工智能技术也随着云计算相关技术的应用而快速发展。因此, 这些技术或研究领域相辅相成, 共同构建了当前蓬勃发展的新一代的信息技术。

本文研究的是面向数据的云计算相关领域, 该领域目前主要分为集中式云计算服务^[2]和边缘云计算服务^[3-4], 本文针对的是集中式云计算服务。集中云计算服务亦被称为云数据中心, 其是信息系统的核心, 也是未来数据

中心建设和发展的方向, 因此, 综述该方向的研究进展具有重要意义, 也可以指导相关科研与实践工作。

关于云计算的概念纷繁多杂, 本文首先区分云计算、云平台、云服务等概念, 具体解释如下:

(1) 云计算是指采用虚拟化等技术构建的数据中心, 然后按使用量付费的模式, 为用户提供各种服务形式的应用。

(2) 云服务是指利用云计算技术, 能够根据用户需求, 提供多元化的服务, 该服务是可扩展、可伸缩的。

(3) 云平台将虚拟化的计算资源、存储资源、网络资源统一管理, 并面向用户提供服务, 形成了云服务。

因此, 这几个概念类而不同, 又相辅相成, 而目前大部分的技术研发、研究机构或公司均是围绕这三个方面进行的。

对于云计算的概念, 目前还存在多种定义, 一般而

综述与评论

Review and Comment

言,云计算是一种按使用量付费的模式,该模式提供可用的、便捷的、按需的网络访问,进入可配置的共享资源池,例如计算、网络、存储等资源,而用户只关心云平台所提供的服务^[5-6]。因此,云计算既指一种可以根据需要动态地提供配置以及取得供应的计算和存储平台,又指一种可以通过互联网进行服务的应用类型^[7]。并且用户的系统规模变化时,云计算系统能够根据用户的需求自由伸缩。

目前,多个领域均涉及云数据中心,本文基于此,进行了大量的对比研究,梳理该场景下云计算的发展现状及关键技术,并提出了下一步的应用方向,为云计算相关研究提供一些参考。

1 云计算简介

本节主要介绍云计算的基本概念、主要特征、体系结构、关注的重点等。

1.1 基本概念及演化过程

2006年谷歌推出“Google 101计划”,并提出了云计算的概念^[2],其提出的基础是随着数据资源的快速增加,需要相应的技术手段筛选或分析其中有意义的信息^[8]。

而陆平等^[9]和钱琼芬等^[10]归纳总结了云计算的演化过程,本文将其进一步细化与延伸,如图1所示。具体而言:

(1)资源汇集阶段:将分散的数据资源逐步实现集中化和标准化,并形成规模化的数据中心。这些中心具有一定的基础设施,为后续数据的有效利用打下基础。

(2)资源应用阶段:为了降低运行成本、缩短业务执行周期、提高资源利用的灵活性,开始出现了物理资源虚拟化的应用;其本质是利用虚拟化技术屏蔽底层物理资源的异构性,进而实现资源利用率的提升和灵活快速的部署业务。

(3)资源共享协作阶段:由于数据中心的各种系统的初始投入较大,软硬件技术则将面临不断升级的压力,从而造成困境,即基于虚拟化的IT架构难以解决不断增加的业务对资源变化的需求。因此,云计算架构开始展现优势,其能够满足业务弹性拓展、按需服务等需求,最终形成一系列的云架构的IT服务。

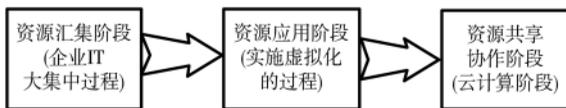


图1 云计算形成过程

1.2 体系结构

所谓云计算,简单地说就是以虚拟化技术为基础,以网络为载体,以用户为主体,并提供基础架构、平台、软件等多种服务形式,整合大规模可扩展的计算、存储、数据、应用等分布式计算资源进行协同工作的超级计算服务模式。其典型体系结构如图2所示。

《电子技术应用》2020年第46卷第8期

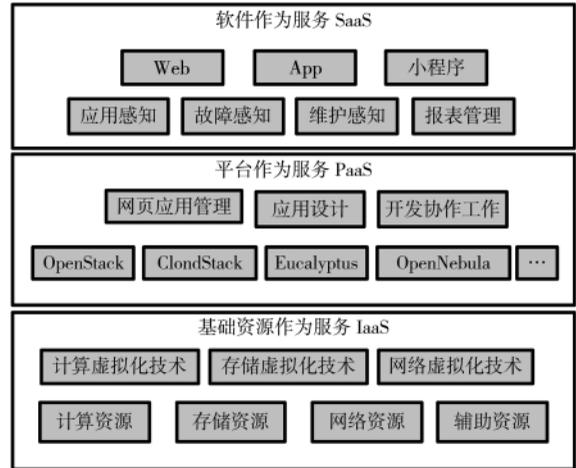


图2 云计算的体系架构

如图2所示,云计算的体系结构通常分为基础设施即服务(Infrastructure as a Service, IaaS)、平台即服务(Platform as a Service, PaaS)和软件即服务(Software as a Service, SaaS)^[11-12]。

其中, IaaS层主要包括计算资源、存储资源、网络资源、辅助资源等设备设施,并据此研发出虚拟化层,即计算虚拟化、存储虚拟化和网络虚拟化。

PaaS层本质是一种云计算服务,能够为企业和个人提供一个应用开发平台或者运行平台,并以SaaS的模式提交给用户。底层硬件和应用基础设施是基于IaaS层实现的,用户可完全不考虑,全心投入其所关注的业务研发工作。

SaaS层主要是面向用户的,可以实现应用感知、故障感知、维护感知和报表管理等多项功能,其实现的形式大多为Web浏览器的形式,最后新兴的也有App程序、小程序等形式。

三者之间的对比关系,如图3所示。



图3 IaaS、PaaS、SaaS层次的对比

1.3 代表性云平台

随着技术或使用模式的发展,企业所考虑的问题不再是将企业的数据和系统搬上云端,而是考虑运用以何种模式构建企业的云。其中的区别在于前者强调必要性,而后者强调的是顶层设计和部署统筹。

综述与评论 Review and Comment

目前,已有的云计算平台主要包括公有云、私有云,或者按照使用类型分为 IaaS、PaaS、SaaS 各类云。本文主要介绍具有代表性的公有云平台 and 私有云平台。结合近年的代表性的云平台,汇总如表 1 所示。其中,阿里作为国内互联网公司的代表,其研发的阿里云计算平台已经涵盖新零售、金融、数字政府、交通物流等行业。

根据国内的信息行业的蓬勃发展,一些行业对于数据安全可靠、运营灵活自由的云计算需求日益凸显,因此私有云应运而生,一些典型的代表如表 2 所示。

2 云计算关键技术的研究

针对数据中心而言,可以分为计算密集型、数据密集型、业务密集型、业务轻量型等,本文主要论述面向数据密集型的数据中心。相对于传统的基础架构,云计算架构通过资源虚拟化和智能化管理,使得应用系统共享基础资源池,并且通过云平台系统的自动化管理,实现快速部署、易于扩展、智能管理。

另外,云计算是分布式计算、并行处理和网格计算的进一步发展,其与并行计算的主要区别在于解决目标任务的方式不同。具体而言,并行计算是整合计算资源快速解决单个的复杂运算问题,而云计算是将各类 IT 资源虚拟化为资源池为各种应用提供按需服务^[13]。因此,云计算是面向数据密集型的,而并行计算是面向计算密集型的。

本节主要介绍云计算系统实现的几个典型关键技术,然而这些技术正在快速发展,本文仅涉及最基本的概念和技术实现。

2.1 虚拟化技术

虚拟化技术是云计算实现的最关键技术之一。通过虚拟化技术,将底层硬件资源资源池化,支持多虚拟机的运行和使用,形成了灵活可扩展的平台服务,提高了

资源利用率。

一般地,虚拟化技术主要涉及计算虚拟化、存储虚拟化和网络虚拟化。

计算虚拟化主要分为 Hypervisor 虚拟化和容器虚拟化。Hypervisor 在虚拟机和底层硬件之间建立一个抽象层,它可以拦截操作系统对硬件的调用,为驻留在其上的操作系统提供虚拟的 CPU 和内存^[14-15]。其中 Hypervisor 虚拟化主要是基于 VMware ESXi、Xen、KVM 等虚拟化技术实现的,这三者的区别如图 4 所示。

	VMware ESXi	Xen	KVM
虚拟化技术	采用全虚拟化(特权指令使用二进制翻译)	采用半虚拟化(特权指令通过 Hypercall 实现)	主要采用硬件辅助虚拟化(全虚拟化的一种)
客户 OS 兼容性	无需修改,兼容性强	需修改客户 OS,兼容性差	无需修改,兼容性强
性能	性能高	I/O 路径太长,性能较差	I/O 路径短,性能高
应用情况及特点	VMware 闭源生态	Citrix 开源生态	Redhat 开源生态

图 4 计算虚拟化技术的对比

而容器虚拟化的性能较快、并发性好,但安全性较差,不适合公有云和多租户场景^[9]。容器技术是一种轻量级的虚拟化技术,用户通过容器镜像来交付应用,其中包含了应用程序及所需的运行时依赖。

存储虚拟化是对存储硬件资源进行抽象化,构成虚拟磁盘系统,进而形成存储资源池,并通过分布式文件管理系统实现业务的提供^[16]。简而言之,用户与底层存

表 1 代表性的公有云计算平台

	Amazon EC2	Google App Engine	Microsoft Azure	Google Documents	Salesforce	阿里云 IaaS/PaaS
类型	IaaS	PaaS	PaaS	SaaS	SaaS	IaaS/PaaS
服务内容	存储、计算、管理和应用服务	程序运行 API 和开发、部署系统平台		Web 应用和服务		资源池/程序托管平台
调用方式	可靠的底层 API 和命令行工具	Web API 和命令行工具		浏览器		Web API
平台	Linux, Windows	Linux, Windows		Linux, Windows		Linux, Windows
特征	提供用户资源虚拟化等基础设施管理及框架	提供在云环境下的程序应用开发运行平台		提供任何时间、地点的应用程序使用	以客户订阅为核心商业模式	提供基础设施租用服务,提供程序开发运行平台

表 2 代表性的私有云计算平台

	阿里云	华为云	浪潮云	麒麟云	其他
类型	私有云	私有云	私有云	私有云	私有云
采用虚拟化技术	KVM 虚拟化技术	Xen 虚拟化技术	KVM 虚拟化技术	KVM 虚拟化技术	KVM 虚拟化技术
运行 OS 环境	Linux	Linux	Linux	Linux	Linux
具体特性	自主率高,性能优异		功能多样,拓展性好		针对特定领域
基础架构	自主研发		基于 OpenStack		多样化

综述与评论 Review and Comment

储资源隔离,在增加使用便捷性的同时,保持相应的性能。存储虚拟化的典型例子就是:每个单位都有办公用品管理部门,取用物品时,只需和管理员沟通,而不用关心具体物品放在哪个库房。

网络虚拟化主要是网络功能虚拟化(Network Function Virtualization, NFV)^[17]、软件定义网络(Software Defined Network, SDN)^[18]。其中, NFV 侧重于网络功能单元的虚拟化,目的是减少资源的支出;SDN 侧重于网络的集中控制、虚拟化和开放接口。两者之间高度互补、相互促进。

值得注意的是,SDN 技术是一种新型的控制与转发分离并直接可编程的网络架构,将数据转发平面和控制平面分离,使得底层设备仅具备数据交换功能,而网络的控制功能集中到 SDN 控制器来实现。因此,SDN 特点是抽象化网络设备,屏蔽底层设备的物理细节差异,并向上层提供统一的管理和 API^[19]。

2.2 分布式编程模型

为处理大规模数据和实现分布式计算,目前使用的主流技术是 Hadoop 项目中的 MapReduce 分布式编程模型。MapReduce 模型是一个简化的并行编程计算模型,它采用函数式编程中的函数来实现并行编程,其核心操作是 Map 和 Reduce^[20]。其每个阶段都将一系列 key/value 对作为输入和输出,其中的键和值的类型由 MapReduce 用户指定。用户同时指定两个函数:Map 函数和 Reduce 函数。Map 操作对每个元素进行操作并生成一组新数据,原数据保持不变,数据之间不存在相关,因此 Map 操作可以高度并行实现;Reduce 操作对 Map 操作的中间结果进行合并,并得到最后的输出^[5]。具体如图 5 所示。

另外,随着应用该模型越来越多,业务场景越来越复杂,该模型也不断被完善^[21]。

近年来,随着数据中心或业务场景的不同,新的分布式计算模型也不断涌现,例如以数据流图 DFG(Data Flow Graph)为基础开发的 TensorFlow 平台^[22]、面向效率提升开发的 Spark 架构^[23]、支持流计算的 Apache Flink 架构^[24-25]等。

Spark 特点是基于内存进行计算且提出了弹性分布式数据集 RDD(Resilient Distributed Dataset),其可以多迭代批量处理数据,还可以兼顾数据库、流处理、图计算等多种范式,是大数据系统领域的全栈计算平台^[26]。而 Flink 是一个可扩展的批处理和流式数据处理的分布式处理引擎。Flink 提供了更多高层的转换算子(如 map/flatMap/filter 等),并通过引入虚节点的概念,Flink 支持算子级并行度调整,并能够将迭代流转换成有向无环图,对需要大量迭代的算法提供支持。

舒娜等^[27]对主流平台 Spark、MXNet、Petuum、TensorFlow 及 PyTorch 等进行对比,并从各角度进行总结。基于此,本文总结如下:这些分布式编程模型或平台均是利用分布式的架构特点、运行某些计算方法,进而解决一些特定场景的问题。不同之处在于各平台提出的分布式计算模型不同:MapReduce 为基本的 Map/Reduce 式的计算模型,而 Spark 定义了一套 RDD 模型,本质上是一系列的 Map/Reduce 组成的一个有向无环图(Directed Acyclic Graph, DAG)。TensorFlow 的计算模型是一张更复杂的图,因为其需要为图中的每个节点和边作出定义。TensorFlow 的这种具体化的定义使它比较适合处理特定类型的计算。

2.3 数据存储和管理技术

该技术由海量数据分布存储技术和数据管理技术组成。

2.3.1 海量数据分布存储技术

云计算系统采用分布式存储,用冗余存储的方式保证数据的可靠性。云计算系统中广泛使用的数据存储系统是 Hadoop 团队开发的 GFS 的开源实现 HDFS(Hadoop Distributed File System)。该文件系统具有高容错性和高扩展性,设计用来部署在低廉硬件组成的集群上,而且可提供高吞吐量来访问应用程序的数据,还提供流式数据访问模式读写超大文件,适合那些有着海量数据集的应用程序^[28]。

Hadoop 文件系统是主从架构,一个 Hadoop 文件系

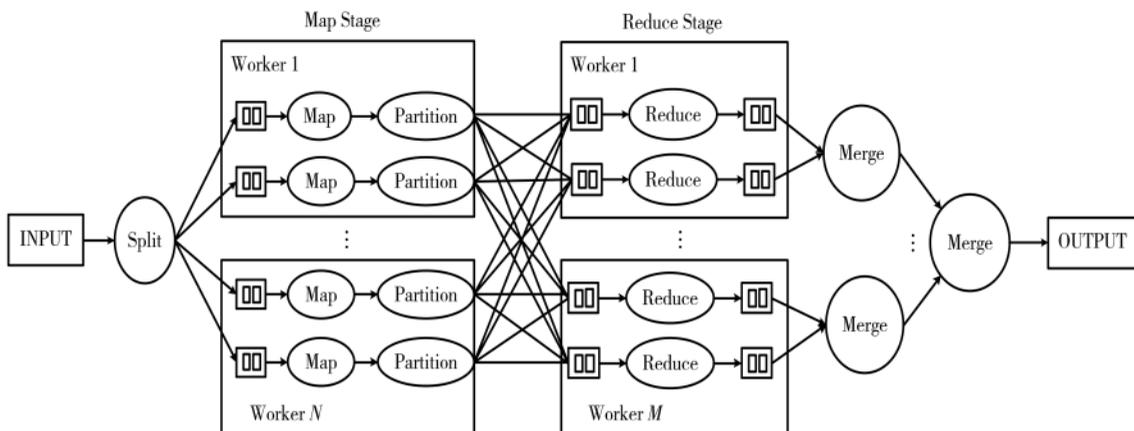


图 5 MapReduce 模型示意图

综述与评论 Review and Comment

统由唯一的一个目录节点(Namenode)和多个数据节点(Datanode)组成,如图6所示。Hadoop文件系统对外表现为一个普通的文件系统,用户(Client)可以用文件名去存储和访问文件,而实际上文件是被分成不同的数据块(Block),这些数据块都存储在数据节点上面^[29-30]。

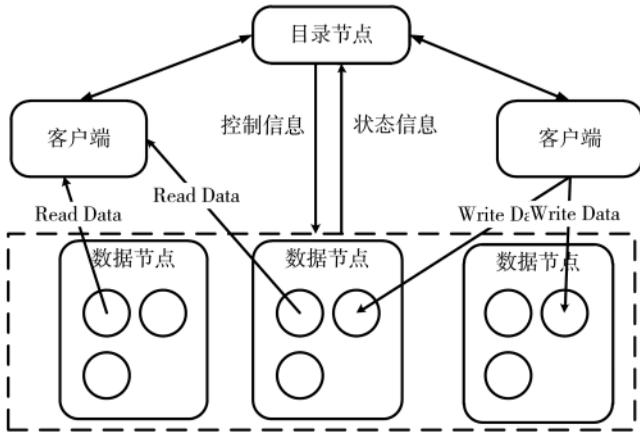


图6 HDFS体系结构示意图

目录节点是集群里面的主节点,负责文件名的维护管理,也是客户端访问文件的入口。数据节点一般就是集群里面的一台机器,负责数据的存储和读取。在写入时,由目录节点分配数据块的保存地址,然后客户端直接写到对应的数据节点。在读取时,当客户端从目录节点获得数据块的映射关系后,就会直接到对应的数据节点读取数据。

2.3.2 数据管理技术

数据管理技术必须能够高效地管理大量的数据,通常数据规模达TB甚至PB级。近年来,云计算系统中的数据管理技术主要是Google的BigTable数据管理技术,以及Hadoop团队开发的开源数据管理模块HBase和Hive,主要用于存储和处理海量结构化数据。

BigTable是一个分布式的结构化数据存储系统,它被设计用来处理海量数据,其将所有数据都作为对象来处理。BigTable是一个稀疏的、多维的和排序的Map,每个Cell由行关键字、列关键字和时间戳来进行三维定位^[31]。

一般地,云计算数据管理层次分为四层:数据组织与管理、数据集成与管理、分布式并行处理和数据分析。本文涉及的为狭义的数据管理技术,其处于数据集成与管理这一层。数据管理的最终目标是实现对非确定性数据的管理与集成,为用户提供高效的查询等服务。例如,在视频监控、智能安防等领域的后台数据中心,可采用BigTable架构来实现。另外,许多学者也基于此进行了大量研究,例如,鲁鹏凯等^[32]提出了RStore大数据存储系统,该系统支持快速应用开发所需的关系数据编程模型,并自动将关系型数据存入BigTable中,同时满足了快速应用开发和高效数据访问两个目标。

2.4 云计算平台管理技术

云平台管理技术能够成为关键技术之一,主要有如下几个方面的原因:

(1)如引言所述,云管理平台将虚拟化的计算资源、存储资源、网络资源统一管理,并面向用户提供服务,形成了云服务,因此,该技术是云计算蓬勃发展的重要支撑。

(2)面向规模庞大的数据中心资源,如何有效管理管理服务器、磁盘阵列等基础设施,并保证系统可靠稳定运行,因此,云平台作为指挥中心,通过智能化管理运维实现业务部署、工作协同和恢复系统故障。

(3)目前主流平台均是基于开源软件平台的,也正是这些平台使得云计算、云服务快速发展,云计算的概念更加深入人心,例如OpenStack。

以OpenStack为例进行说明。作为一种典型的开源云计算管理平台项目,其目标是提供实施简单、可大规模扩展、标准统一的云计算平台^[9]。OpenStack包括网络、操作系统、服务器等方面,适用于各个领域的应用。同时,其是一系列组件服务集合,具体包括Horizon、Nova、Neutron、Glance、Cinder、Swift等组件。目前OpenStack正处于高速发展阶段,世界级供应商都是其支持者。

3 基于数据类型的云计算应用

随着业务的发展,越来越多的企业需要多个数据中心,同时生成大量的数据,而这些数据可能是来自不同硬件环境、不同IT管理架构、面向多样化的业务需求的,因此,考虑以何种模式构建云系统是至关重要的。

目前,一种典型的云计算系统分类是私有云、公有云、混合云。然而,云数据的核心是数据,数据种类是整个领域的源头,因此,本文将数据分为三类:消费应用类数据(商业云)、工业应用类数据(工业云)、特种行业应用类数据(特种行业云)。据此,从行业应用的角度,本文对近年来云服务的应用情况进行综述。

3.1 商业云

作为最开始应用云计算技术的领域,各类商用云是由行业内的主流公司发起的,它们也致力于该方面的研发与应用。具体包括Google的Google App Engine、IBM的RC2、Amazon的弹性计算云EC2、Microsoft的Azure,以及国内的华为云、腾讯云、百度云、阿里云等。

这些云均属于公有云范畴,其典型的应用是用于构建大型数据中心,形成消费类数据,进而用于完成各类搜索、电商、社交等业务领域的功能需求。消费类数据包括结构化数据、非结构化数据和其他多样化数据,因此,这类云计算应用的范围更广,对技术研发适应度更高。以阿里云为例,其依托强大的技术创新能力,可提供云原生、数据智能、零售、金融、制造等多个领域的解决方案。

一般地,商用云的数据中心系统架构一般如图7所示,主要由基础设施、数据中台(支撑服务)、综合呈现三层构成。许多企业是通过搭建标准化基础设施,实现资



图7 典型的商用云系统架构

源的共享,并在基础上研发自身业务领域的服务或平台。其中,数据中台主要用于实现数据集成、数据洞察和平台化,而数据洞察更侧重于提供一个数据分析平台,为各类业务提供智能化筛选和分析处理框架和解决方案^[12]。

关于业务应用,主要分为提供云资源的云平台供应商和基于云平台研发应用的服务供应商。例如,作为第一梯队的阿里、腾讯、百度、滴滴、美团、字节跳动等公司既是云平台供应商,也是服务供应商。

另外,随着社会发展的需要,远程协作办公也逐渐兴起,例如钉钉和企业微信是依托腾讯云和阿里云推出的在线协作软件产品。2020年年初,为应对新型冠状病毒的疫情,确保网络通畅和用户体验,阿里云技术人员紧急扩容网络和计算资源,利用弹性计算资源编排在2小时内新增10 000余台云服务器。

因此,商业云促进了互联网业务的快速发展,并基于此不断涌现新的行业或领域;同时,各类互联网业务的应用,也促进了云计算相关技术的不断完善。

3.2 工业云

不同于其他云计算系统,工业云是基于云计算架构,并结合工业生产和应用,收集、存储和分析工业数据,提供的云服务,涉及产品研发设计、实验和仿真、工艺设计、加工制造等环节^[33]。

而物联网相关技术的发展,许多公司开始构建基于物联网的工业云平台,例如GE的PaaS软件平台Predix、西门子的MindSphere、三一重工企业控制中心(Enterprise Control Center, ECC)客户服务平台、徐工与阿里携手搭建的“徐工工业云”^[34]。

目前,工业云平台主要是将常用的工业软件迁移到云上,以在线服务的形式向用户提供CAD、CAE、PLM等工业软件服务,具体架构如图8所示。工业云逻辑架构主要包括物联网感知层、数据传输层、云服务层、应用服务层。其主要利用云服务的相关技术,用于实现面向不

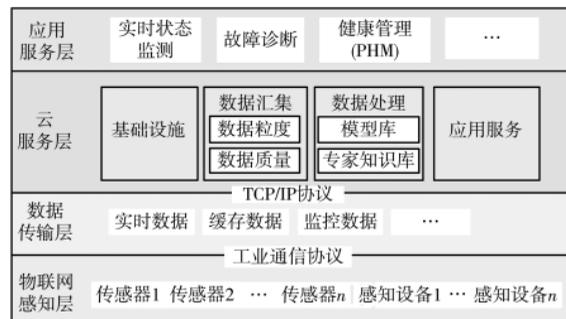


图8 典型的工业云逻辑架构

同工业领域的智能化服务,例如,面向流程工业的实时状态监测、面向动力设备的故障诊断^[34]、面向选矿设备的健康管理^[35]等。

工业云与商业云不同之处主要有两个方面:(1)数据传输复杂:数据传输层需考虑通信协议转换网关,将工业标准数据转换为IP数据,进而利用云服务层中的云计算技术实现上层的领域业务实现。(2)应用对象不同:工业云面向工业过程,其应用场景相对固定,实时性和可靠性要求高(否则引起严重的大范围事故),而商业云面向商业用户或个人用户,场景多样化,数据多样化,需求灵活多变,但实时性和可靠性相对低一些,更适合云计算系统的应用及发展思路。

3.3 特种行业云

除了上述两类云计算系统,还有一些针对特殊行业的云计算系统,例如国防、政务、公安等。本文主要针对国防领域进行简要介绍。

特种行业云通常为私有云,即针对某类应用搭建用于特定用途的云计算系统。其典型应用是构建大型数据中心,用于信息的采集、数据的处理、指挥控制和综合态势展现等。特种行业云的典型逻辑架构如图9所示。

该类云主要由基础层、服务层、应用层组成,其中基础层和服务层大多是基于OpenStack平台研发的,虚拟



图9 典型的特种行业云逻辑架构

化原理和资源共享模式也是类似的。与其他类型的云不同的是,特种行业云更注重数据的安全和整个系统的可靠。这个数据可能大部分均为结构化类型的数据,易于处理,但是实时性和准确性要求更高。

针对该领域,已有一些学者进行了深入的调研和研究,文献[36]针对军事领域综述了如何应用云计算技术;文献[37]研究了美军云计算安全策略方面的最新进展,其指出可借鉴美军云计算安全策略的良好经验,采取渐进式发展云计算安全策略;文献[38]针对海上战术环境平台,研究了战术云环境下信息总线构建和信息服务高可用集群等关键技术,并对典型组织应用场景进行演示。

4 结论

云计算是传统数据分析(数据挖掘、统计分析、机器学习、智能算法)的升级,也是虚拟化技术、分布式计算、海量数据存储与管理技术的融合发展。本文从数据角度出发,首先介绍了云计算的概念、体系结构和代表性云平台,然后综述了当前云计算领域相关的关键技术,最后从应用角度,提出了运用数据来源来分析云计算应用的方法,将其分为商业云、工业云、特种行业云,并重点介绍了相应架构和应用特点。

尽管随着信息技术发展,云计算的研究和应用越来越全面,但仍然有几个方面需要重点考虑:(1)业务种类繁多,需要考虑搭建异构云平台,最大程度实现资源共享;(2)新的应用方面不断涌现,云计算架构和关键技术需要不断迭代升级,例如边缘计算;(3)云计算在中国的应用潜力巨大,但同时要考虑云安全的问题,在不影响应用的前提下,国内云计算供应商需要研制针对中国用户的安全稳定版本,确保核心技术实现国产化。

参考文献

- [1] 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),2014,48(6):957-972.
- [2] 张建勋,古志民,郑超.云计算研究进展综述[J].计算机应用研究,2010,27(2):429-433.
- [3] 施巍松,孙辉,曹杰,等.边缘计算:万物互联时代新型计算模型[J].计算机研究与发展,2017,54(5):907-924.
- [4] YAQOUB I, AHMED E, GANI A, et al. Mobile ad hoc cloud:

A survey[J].Wireless Communication & Mobile Computing, 2016, 16(16):2572-2589.

- [5] 张建华,吴恒,张文博.云计算核心技术研究综述[J].小型微型计算机系统,2013,34(11):2417-2424.
- [6] 蒋永生,彭俊杰,张武.云计算及云计算实施标准综述与探索[J].上海大学学报(自然科学版),2013,19(1):5-13.
- [7] 何嘉,彭商濂.云数据中心虚拟机管理研究综述[J].电子科技大学学报,2016,45(1):107-112.
- [8] CHEN K, ZHENG W. Cloud computing: system instances and current research[J]. Journal of Software, 2009, 20(5): 1337-1348.
- [9] 陆平,赵培,王志坤,等.云计算基础架构及关键应用[M].北京:机械工业出版社,2016.
- [10] 钱琼芬,李春林,张小庆,等.云数据中心虚拟资源管理研究综述[J].计算机应用研究,2012,29(7):2411-2415.
- [11] 李乔,郑啸.云计算研究现状综述[J].计算机科学,2011,38(4):32-37.
- [12] ARMBRUST M, FOX A, GRIFFITH R, et al. A view of cloud computing[J]. Communications of the ACM, 2010, 53(4):50-58.
- [13] 邹复民,蒋新华,胡惠淳,等.云计算研究与应用现状综述[J].福建工程学院学报,2013,11(3):231-242.
- [14] 余时强,张为华.GPU虚拟化相关技术研究综述[J].计算机系统应用,2017,26(12):25-31.
- [15] ANDR H, LAGAR-CAVILLA S, WHITNEY J A, et al. Snowlock: rapid virtual machine cloning for cloud computing[C]. Proceedings of the 4th ACM European Conference on Computer Systems, Nurem-berg, Germany, 2009: 1-12.
- [16] Xia Qichao, Jia Lianxing. Reserch of integration military storage resources based on storage virtualization[C]. 2006 International Workshop on Networking, Architecture, and Storage, 2006.
- [17] 王进文,张晓丽,李琦,等.网络功能虚拟化技术研究进展[J].计算机学报,2019,42(2):185-206.
- [18] KREUTZ D, RAMOS F M V, VERISSIMO P E, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2015, 103(1): 14-76.

综述与评论 Review and Comment

[19] CESELLI A, PREMOLI M, SECCI S. Cloudlet network design optimization[C]. Proceedings of the IFIP Networking, Toulouse, France, 2015: 1-9.

[20] DEAN J, GHEMAWAT S. Map Reduce: simplified data processing on large clusters[C]. Proceedings of the 6th Conference on Symposium on Operating Systems Design & Implementation, 2004.

[21] 夏靖波, 韦泽鲲, 付凯, 等. 云计算中 Hadoop 技术研究与综述[J]. 计算机科学, 2016, 43(11): 6-11.

[22] ABADI M, AGARWAL A, BARHAM P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems[J]. arXiv: 1603.04467, 2016.

[23] ZAHARIA M. An architecture for fast and general data processing on large clusters[M]. Morgan & Claypool, 2016.

[24] CARBONE P, KATSIFODIMOS A, EWEN S, et al. Apache flink: stream and batch processing in a single engine[J]. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2015, 36(4).

[25] 马黎. Spark 和 Flink 平台大数据批量处理的性能分析[J]. 中国电子科学研究院学报, 2018, 13(2): 191-195.

[26] 廖湖声, 黄珊珊, 徐俊刚, 等. Spark 性能优化技术研究综述[J]. 计算机科学, 2018, 45(7): 7-15.

[27] 舒娜, 刘波, 林伟伟, 等. 分布式机器学习平台与算法综述[J]. 计算机科学, 2019, 46(3): 9-18.

[28] 方巍, 文学志, 潘吴斌, 等. 云计算: 概念、技术及应用研究综述[J]. 南京信息工程大学学报(自然科学版), 2012, 4(4): 351-361.

[29] 张栗粽, 崔园, 罗光春, 等. 面向大数据分布式存储的动态负载均衡算法[J]. 计算机科学, 2017, 44(5): 178-183.

[30] BYRNE J, CASARI P, EARDLEY P, et al. Reliable capacity

provisioning for distributed cloud/edge/fog computing applications[C]. Proceedings of the 26th European Conference on Networks and Communications. IEEE, 2017: 1-6.

[31] 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.

[32] 鲁鹏凯, 江大伟, 陈珂, 等. RStore: 基于 Big Table 的关系数据模型存储系统[J]. 计算机科学与探索, 2018, 12(10): 1547-1558.

[33] 徐泉, 王良勇, 刘长鑫. 工业云应用于技术综述[J]. 计算机集成制造系统, 2018, 24(8): 1887-1901.

[34] 杨文广, 龙泉, 蒋东翔. 基于云平台的动力设备远程故障诊断系统研究[J]. 制造业自动化, 2017, 39(7): 6-14.

[35] 侯一鸣, 徐泉, 李亚杰, 等. 基于物联网和工业云的选矿设备状态监控系统[J]. 计算机集成制造系统, 2017, 23(9): 1972-1982.

[36] 陈卫, 杨长虹, 戴超凡. 云计算在军事领域的应用[J]. 国防科技, 2015, 36(4): 23-26.

[37] 程妍妍. 美军云计算安全策略发展及启示[J]. 国防科技, 2019, 40(1): 51-57.

[38] 唐素纯, 李宁, 于钺, 等. 面向海上战术云的信息资源服务架构设计[J]. 舰船电子工程, 2019, 39(2): 103-109.

(收稿日期: 2020-04-15)

作者简介:

葛文双(1989-), 通信作者, 男, 硕士研究生, 工程师, 主要研究方向: 数据驱动的智能评估、云计算与数据分析, E-mail: wsedianqi@126.com.

郑和芳(1985-), 男, 硕士研究生, 工程师, 主要研究方向: 计算机应用技术。

刘天龙(1988-), 男, 硕士研究生, 工程师, 主要研究方向: 指挥控制系统总体技术。

Physics, 2020, 118(5): 483-524.

[2] IEC TR 62669. Case studies supporting IEC 62232-determination of RF field strength and SAR in the vicinity of radiocommunication base stations for the purpose of evaluating human exposure[S]. 2018.

[3] ITU-T Recommendation K.100. Measurement of radio frequency electromagnetic fields to determine compliance with human exposure limits when a base station is put into service[S]. 2018.

[4] ITU-T Recommendation K.Sup16. Electromagnetic field compliance assessments for 5G wireless networks[S]. 2018.

(收稿日期: 2020-06-11)

作者简介:

谈儒猛(1973-), 男, 硕士, 工程师, 主要研究方向: 5G 辐射、节能、电磁兼容。

熊尚坤(1977-), 男, 博士, 高级工程师, 主要研究方向: 5G 移动通信、无线网络演进。

武彤(1975-), 男, 硕士, 副研究员, 主要研究方向: 节能计量、检测。

(上接第 45 页)

天线波束的随机性、用户在时间和空间上的分布、实际 EIRP 的确定等)。5G 由于引入大发射功率和大规模 MIMO 天线技术, 使得理论评估变得复杂, 传统的理论评估方法将得出过于保守的安全边界。鉴于目前 5G 网络尚处于商用初期, 用户数量和用户规模亟需培育发展, 因此当前阶段难以开展真实用户场景下的现场测量评估。目前正在起草的行业标准中推荐采用模拟加载(如 Speedtest 软件或下行灌包等)方式进行现场测量, 这种人为加载方式测得的环境电磁场并不能反映 5G 网络带来的真实电磁环境分布, 仅作为评估 5G 基站在严酷场景下的电磁辐射水平情况。采用本文介绍的理论与评估方法基本可以符合未来 5G 网络实际, 文中给出的理论评估算法、实际 EIRP 确定和理论评估考量因素均应进一步验证与探索, 使理论评估更加符合实际电磁场的分布。

参考文献

[1] ICNIRP(2020). ICNIRP guidelines for limiting exposure to electromagnetic fields(100 kHz to 300 GHz)[J]. Health

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所