基于普适云的大数据挖掘①

时念云, 王文佳, 马 力

(中国石油大学(华东) 计算机与通信工程学院, 青岛 266580)

摘 要: 通过查阅大量相关文献资料, 概述大数据和普适云的基本概念, 提出基于普适云的大数据挖掘架构, 在理论方面论证其可行性、论述其运行模式并进行了性能分析, 分析总结了基于普适云的大数据挖掘所涉及的关键技术.

关键词: 普适计算; 云计算; 大数据; 数据挖掘; UbiCloud

Data Mining of Big Data Based on UbiCloud

SHI Nian-Yun,WANG Wen-Jia, MA Li

(College of Computer and Communication Engineering, China University of Petroleum(Huadong), Qingdao 266580, China)

Abstract: By accessing to a large number of relevant documents, this paper outlined the basic concepts of Big Data and UbiCloud, and proposed data mining architecture of Big Data based on UbiCloud. Then, this thesis demonstrated the feasibility of the architecture, discussed the operating mode of it and analyzed its performance, in theory. Moreover, it analyzed and summarized the key technology involved with data mining of Big Data based on UbiCloud.

Key words: ubiquitous computing; cloud computing; big data; data mining; ubiCloud

1 引言

随着物联网与移动通信等相关技术的发展, 大数据已成为继“云计算”之后的又一信息科技新热点. 2012 年 3 月 29 日, 美国 政府拨款 2 亿美元启动“大数据研究和发展倡议”计划, 标志大数据已成为科技信息关注的重点[1].

目前, 大数据尚未有明确的定义, 维基百科中关于大数据的定义描述: 在信息技术中, “大数据”是指一些使用目前现有数据库管理工具或传统数据处理应用很难处理的大型而复杂的数据集. “大数据”研究机构 Gartner 给出的定义则为: “大数据(Big data)”是指需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产. 关于大数据的定义, 虽然尚未获得统一的认可, 但大都认为大数据应该具有 4 个特点: 规模性(Volume), 高速性(Velocity), 多样性(Variety)和真实性(Veracity). 业界将其归纳为 4V 定义.

大数据无疑是由信息技术的不断发展催生的. 数据的产生方式大体经历了三个阶段: 运营式系统阶段, 该阶段的数据产生方式是被动的; 用户原创内容阶段, 该阶段的数据产生方式是主动的; 感知式系统阶段, 该阶段的数据产生方式是自动的. 大数据的数据来源正是由这些被动、主动和自动的数据共同构成的, 但其中自动式的数据才是大数据产生的最根本原因[2]. 而普适计算网络正是自动式数据产生主要平台. 普适计算技术是由 Mark Weiser 于 1991 年正式提出的. 它的核心思想是: 小型、便宜、网络化的处理设备广泛分布在日常生活的各个场所, 用户通过这些设备可以在任何时间、任何地点、以任何方式进行信息的获取与处理. 在普适计算环境中, 无线传感器网络将广泛普及, 从而自动地产生海量数据. 同时, 普适环境中, 其他的终端设备也会或被动或主动地产生大量数据. 也就是说, 普适计算网络是“大数据”的主要数据来源平台.

相比于传统的“海量数据”, 大数据虽然具有 4V 特征, 但其在规模上依然是“海量数据”(从 TB 级别, 跃

① 基金项目:中央高校基本科研业务费专项资金(14CX02032A)
收稿时间:2013-04-17;收到修改稿时间:2013-05-20

升到 PB 级别), 为了从中获得潜在的有价值的信息, 我们依然要采用深层次的分析技术如数据挖掘进行知识模式的发掘. 一些研究机构已经进行了相关的实例应用研究. 例如, Global Pulse[3]研究社交网络中情绪和失业率之间的关系发现: 在爱尔兰, 当社交网络上“困惑”和“沮丧”这些指标升高 3 个月后, 失业率也会升高; 在美国发生在失业率升高之前持续升高的是“愤怒”这一指标; 而在“失业”指标上升 2 个月后人们在谈论“房子”, 这也许意味着他们准备卖掉自己的房产; 在过后的几个月, 谈论“公交”和“地铁”的在上升, 这也许意味着他们承担不起开车的油费, 或者已经准备将车卖掉. 这些研究将对政府的工作提供决策支持. 例如, 当“困惑”和“沮丧”这些指标在社交网络上出现升高趋势时, 政府可以预估失业率将上升, 并及时采取相应措施避免大范围失业情况的发生, 从而减小由失业而引发的一系列消极事件发生的概率, 维护良好的社会秩序和经济秩序.

既然大数据有别于传统的“海量数据”, 相应的数据挖掘技术也必需有所改进. 传统的单机或简单集群的分布式挖掘算法, 已不能适应大数据的 4V 特性. 而基于普适云的分布式挖掘算法却可以在一定程度上弥补现有算法的不足. 普适云[4], 顾名思义, 是一种面向普适终端的云计算, 它可以更好地满足普适环境的性能要求. 但它在本质上依然属于云计算的范畴. 云计算是以数据为中心的一种数据密集型的超级计算. 根据美国国家标准与技术研究院(NIST)的定义: 云计算是一种利用互联网实现随时随地、按需、便捷地访问空想资源池(如计算设施、存储设备、应用程序等)的计算模式. 分布式数据挖掘平台的优势正是云计算的本质. 它在数据存储、数据管理、编程模式、并发控制、系统管理等方面具有自身独特的技术: 海量分布式存储技术、并行编程模式、数据管理技术、分布式资源管理技术、云计算平台管理技术. 上述各种技术可以为大数据的数据挖掘提供技术和平台支撑.

2　系统架构与运行模式

普适计算推动了大数据时代的到来, 云计算为大数据时代的知识发现提供了技术支撑与处理平台. 鉴于普适计算与云计算所特有的技术特征以及大数据的 4V 特性, 我们提出了如图 1 所示的, 基于普适云的大数据挖掘模型架构.

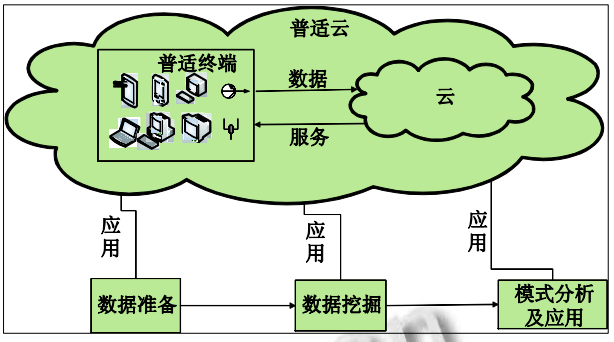

图 1　基于普适云的大数据挖掘模型架构

首先, 该架构与传统的数据挖掘的不同之处在于, 知识发现的过程中少了问题描述这一阶段, 原因在于大数据时代先有数据再有模式, 而传统的知识发掘, 往往是针对一定的知识模式进行相关的数据采集. 其次, 该架构中所涉及的终端形态各异, 因而数据将存在更大的异构性, 包含结构化, 半结构化, 无结构化三大类数据. 以上两点都在一定程度上增加了数据挖掘的难度. 该架构中所涉及的普适云的详细模型如图 2 所示.

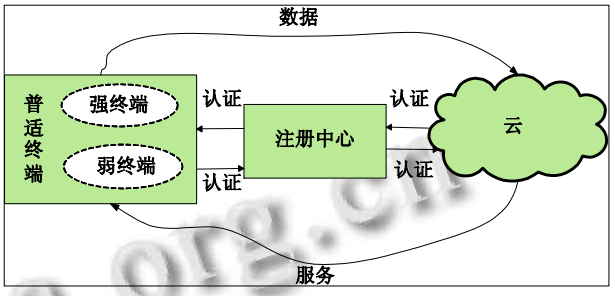

图 2　普适云的详细模型

该模型的运行模式详述如下:

(1) 普适客户端在注册中心通过一定的动态协议如 UPNP 协议进行授权认证, 注册信息涉及终端多方面性能信息如内存、待机时间等.

(2) 云计算服务器在注册中心通过一定的动态协议如 UPNP 协议进行授权认证, 注册信息涉及服务器多方面性能与所能提供的应用服务类型, 如 CPU 性能, Word 文件打开功能等.

(3) 注册中心只为在此注册认证的客户端和云计算资源提供服务, 动态地发现相互适配的客户端和云计算服务.

(4) 客户端动态地向注册中心申请获得或释放云

计算资源; 云计算服务器动态地向注册中心广播 Notify 自己的当前状态(忙碌或闲置)以及运行状态, 使得系统的负载均衡达到最佳状态, 提高运算效率.

(5) 客户端向注册中心发出服务请求, 注册中心通过综合考虑客户端所提供的信息与先前的注册信息, 为客户端提供最佳服务推荐. 具体体现为: 1)针对强终端: 当它能负荷其所涉及的运算时, 则云只向其提供服务, 两者之间不进行数据传输, 这将在一定程度上可以降低通信代价; 当它不能负荷其所涉及的运算时, 客户端需要把数据传输给云计算资源进行处理. 2)针对弱终端, 大部分情况下, 它都不能负荷其所涉及的运算, 需要将数据传输给云计算资源进行处理; 一般情况下, 弱终端也可以申请云服务, 因为陈援非, 崔丽等在文献[4]中通过实验验证知: 当网络畅通时, 使用云服务应用的响应时间比使用本地计算模式的响应时间更低. 这样也就不必在弱终端安装应用软件了, 从而减少弱终端的功耗, 延长待机时间. 3)云计算资源提供两大类数据处理平台, 一个是流处理平台, 另一个是批处理平台. 前者主要是处理一些实时性比较高的数据, 例如消费者购买意图或者点击预估数据. 后者更多的是做一些基础任务, 例如全网行为的挖掘分析、BI 分析、商务报表等, 完成一些非实时性任务. 注册中心则会根据客户端的服务要求, 推荐最佳的云服务资源, 提高数据挖掘效率.

在该模型中, 注册中心可类比为人类的神经中枢, 进行任务的匹配与调度, 而“云”则可类比为人类的器官, 完成注册中心对于终端服务请求做出的响应指令. 即, 大数据挖掘的真正执行者是“云”. 云计算通常采用 Map/Reduce 编程模式, MapReduce 将数据处理任务抽象为一系列的 Map(映射)--Reduce(还原)操作对. Map 主要完成数据的过滤操作, Reduce 主要完成数据的聚集操作. MapReduce 是面向由数千台中低端计算机组成的大规模机群而设计的, 其扩展能力得益于其 shared--nothing 结构、各个节点间的松耦合性和较强的软件级容错能力: 节点可以被任意地从机群中移除, 而几乎不影响现有任务的执行. 但是, 现有的 MapReduce 方法, 如 Hadoop 等属于对持久化数据的静态批处理方式, 实时性受限. 为了改善现有方法的这一缺陷, 文献[5]提出了一种面向大规模历史数据的数据流处理模型 RTMR(real-time MapReduce), 其主要的架构形式有: RTMR(n)(其中, n 属于正整数), RTMR(0). 鉴于普适计算网络的数据流速度受采集端带宽等因素影响, 普适云中的“云”应选用 RTMR(0) 集群构造方式. RTMR(0)的系统架构模式如图 3 所示. 该架构模式在数据规模不断扩大的情况下适用性更强.

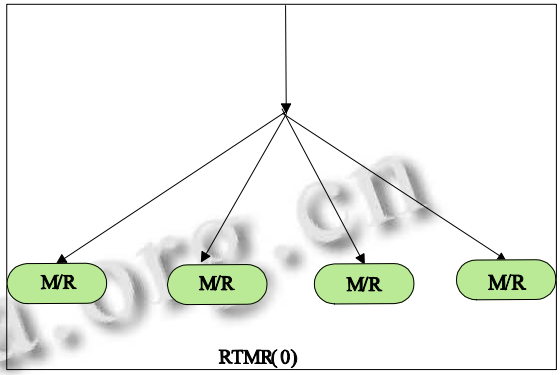

图 3 RTMR(0)系统架构模式

该系统架构的性能分析可以从以下几个主要方面展开: 1)伸缩性. MapReduce 的 shared--nothing 结构、节点间的松耦合性与较强的软件容错能力等扩展特性, 使得节点的移除与增添具有很大的灵活性, 而且节点的这种变更不会影响现有任务的执行. 所以该系统架构具有良好的伸缩性. 2)通信代价. 该系统架构所涉及的通信代价计算公式: 通信代价=数据传输通信代价/应用程序传输通信代价+注册信息交互传输通信代价+数据加密解密信息传输通信代价. 显然, 在大数据情况下, 数据传输通信代价无疑占有最大比例的通信代价. 系统架构中针对强终端所采取的策略: 当它能负荷其所涉及的运算时, 则云只向其提供服务, 两者之间不进行数据传输. 该策略无疑将在一定程度上可以降低总的通信代价, 减轻网络的负担. 3)灵活的安全策略. 通过注册中心的信息控制与策略选择, 可以设定灵活的数据安全策略. 当所进行的相关操作处于内部网络时, 可以适当降低数据加密/解密的复杂度, 从而减少计算的代价, 提高运算时效性; 当所进行的操作与外部网络有交叉时, 可以适当增加数据加密/解密的复杂度, 从而获取更高的数据安全性, 减少隐私泄露. 4)负载均衡控制. 注册中心根据客户端与云计算资源的信息动态地匹配资源与服务, 从而推荐最佳服务, 将使得该架构达到较好的负载均衡状态.

3 关键技术分析

基于普适云的大数据挖掘架构所涉及的关键技术

主要有: 普适云技术与数据挖掘技术.

3.1 普适云技术

普适云主要涉及到云计算技术和普适计算技术. 二者需要通过一定的认证协议与相关技术来架构普适云, 使得普适终端能够方便、快捷、无缝、安全地获得最佳的云资源服务.

3.1.1 云计算技术

“云计算”构想[6,7]是于 2006 年由 Google、Amazon 等公司提出的. 它是网格计算(Grid Computing)、分布式计算(Distributed Computing)、并行计算(Parallel Computing)、效用计算(Utility Computing)、网络存储(Network Storage Technologies)、虚拟化(Virtualization)、负载均衡(Load Balance)等传统计算机技术和网络技术发展融合的产物. 云计算的定义可以从狭义和广义两方面来看. 狭义云计算是一种 IT 基础设施的交付和使用模式, 指通过网络按需、易扩展的方式获得所需的资源(包括硬件、平台和软件). 提供资源的网络就是“云”. “云”中的资源在使用者看来是可以无限扩展的, 并且可以随时获取, 按需使用, 随时扩展, 按使用付费. 这种特性被人们形象地称为像使用水电一样使用 IT 基础设施. 广义云计算是指服务的交付和使用模式, 指通过网络已按需、易扩展的方式获得所需的服务. 这种服务可以是 IT 和软件、互联网相关的, 也可以是任意其它服务. 它意味着计算能力也可作为一种商品通过互联网进行流通.

作为信息产业一大支柱技术, 云计算模式一经提出便得到了工业界和学术界的广泛关注, 使得云计算技术逐步趋于成熟. Hadoop 等新型并行编程框架简化了海量数据处理模型. Amazon 等公司的云计算平台提供可快速部署的虚拟服务器, 实现了基础设施的按需分配. Salesforce 公司的客户关系管理(CRM, customer relationship management)服务等云计算服务将桌面应用程序迁移到互联网, 实现应用程序的泛在访问. Google 公司的 App Engine 云计算开发平台为应用服务提供商开发和部署云计算服务提供接口. VMware 和 RedHat 等为云计算的虚拟化实现提供了技术平台. 同时, 学术界也对云计算展开了大量研究工作. 2007 年, 包括斯坦福大学在内的多所美国高校便开始和 Google、IBM 合作, 研究云计算的关键技术. 此外, 各国政府纷纷也将云计算列为国家战略, 投入了相当大的财力和物力用于云计算的部署. 其中, 美国政府利用云计算技术建立联邦政府网站, 以降低政府信息化运行成本. 英国政府建立国家级云计算平台(G-Cloud), 超过 2/3 的英国企业开始使用云计算服务. 在我国, 北京、上海、深圳、杭州、无锡等城市开展了云计算服务创新发展试点示范工作; 电信、石油石化、交通运输等行业也启动了相应的云计算发展计划, 以促进产业信息化. 当前中国的云计算处于成长期, 预期到 2015 年之后, 中国云计算产业将真正进入成熟期, 云计算服务模式将被更多用户接受. 在中国, 云计算系统集成商主要有东软集团、华为和中兴通讯等; 云计算平台提供商主要有浪潮信息、中国软件和方正科技等; 云计算服务提供商主要有中国移动和中国电信等; 应用开发商主要有焦点科技等. 在中国的高校中, 中国科学院与东南大学等都已经建立了云计算研究中心和应用平台, 展开了对云计算的研究.

3.1.2 普适计算技术

普适计算以间断连接与轻量计算为重要特征, 是一种无所不在随时随地可以进行计算和获取信息服务的新型计算模式, 它强调和环境融为一体的计算, 其根本原理是在信息社会中降低使用设备的复杂程度, 将“人围绕机器转”的现状变革为“机器围绕人转”, 从而使人们的工作学习更为轻松有效, 在真正意义上实现以人为本的生活方式. 普适计算的含义十分广泛, 所涉及的技术包括移动通信技术、小型计算设备制造技术、软件与中间件技术、智能感知技术、人机交互技术等. 普适计算降低了设备使用的复杂性[8]. 它主要针对移动设备, 比如信息家电或某种嵌入式设备, 如掌上电脑、BP 机、车载智能设备、笔记本计算机、手表、智能卡、智能手机(具有掌上电脑的一部分功能)、机顶盒、POS 销售机、屏幕电话(除了普通话机的功能外, 还可以浏览因特网)等新一代智能设备. 普适计算设备可以一直或间断地连接着网络. 与 Internet、Intranet 及 Extranet 连接, 使用户能够随时随地获取相关的各种信息, 并做出回应. 由于普适计算设备的高度移动性, 所以也被称为移动计算. 虽然目前实现真正意义上的普适网络仍存在很多困难, 但随着移动终端、感知智能设备以及相关软件技术的提高, 特别是通信网络、智能设备与无线传感器技术的迅速发展, 普适计算将逐步趋于成熟. 国外一些典型的研究项目[9]主要有: (1)麻省理工学院的 Oxygen 项目. 其寓意是: 未来计算像氧气一样无处不在并可自由获取. 该项目将固定计算设备和移动设备

通过自动配置的网络连接起来. 系统采用了包括休眠环境的自动转换等 8 种环境驱动技术. (2)Microsoft 公司的 Easy Living 研究项目. 致力于智能环境的体系开发, 涉及中间件、几何世界建模、定位感知、服务描述等技术. 其关键特点是: 机器视觉、多传感器的自动和半自动校准以及独立于设备的通信. (3)AT&T 实验室和英国剑桥大学合作的研究项目 Sentient Computing. 通过用户接口、传感器以及建立资源数据等手段, 为系统提供基于用户和位置的数据更新能力, 系统可无缝扩展到整个建筑物. (4)卡内基•梅隆大学的 Aura 项目. 强调普适计算的中间件技术和应用设计, 该项目包括三个子项目: Darwin 智能网络是 Aura 的核心; Coda 分布式文件管理系统; Odyssey 为资源自适应提供操作系统支持. 该系统可容纳桌面、手持和可穿戴系统. 此外, 还有惠普公司的 Cool Town 项目、Everyday Computing 项目; IBM 的 WebSphere Everyplace 项目; 华盛顿大学的 Portolano 项目等等. 目前国内也非常重视普适计算的研究, 并将其列入国家自然科学基金委信息科学部 2003 年资助的 18 个重点项目之一.

3.1.3 普适云的架构

普适云的架构研究已经取得了一定的成果. 陈援非, 崔丽等构建了在 UPNP 协议之上利用 SoA 技术进行计算资源发布的异构计算系统, 该系统可以为普适终端提供无缝的云资源调用接口, 从而降低终端自身对资源的需求. Huifeng[10]提出了一种高性能的远程计算平台, 使得瘦客户端可以远程访问服务器的应用程序. 牟权, 叶保留与陆桑璐[11]提出了一种基于云计算的普适服务集成平台. 童晓渝, 张云勇与徐雷[12]以云计算为核心, 以普适计算为触角, 提出了智能普适网络的系统架构, 支撑移动互联网、物联网和“三网融合”的应用. 普适云技术的成熟是大数据挖掘架构实现的关键. 普适计算技术的发展缓慢则是限制普适云架构的重要的技术瓶颈之一. 随着小型计算设备制造技术、软件与中间件技术与嵌入式技术等技术的逐步发展, 普适终端的各方面性能已有所改善. 但是, 普适环境中普遍存在的一些弱终端, 如无线传感器、智能手机和瘦客户机等, 它们与服务器相比较, 在内存、磁盘容量、待机时间和处理速度方面依然处于劣势. 因而限制了一些应用程序的运行实施, 增加了普适云架构的实现难度. 另外, 普适终端生产技术良莠不齐的发展与某些生产标准的不一致性, 促使普适终端的性能存在很大的差异, 以至于云计算服务很难满足通用性, 降低了普适云的适用性与实用性. 普适终端制造技术的发展与通用性算法的实现已然成为亟需解决的问题. 此外, 无线网络技术的发展也是推动普适云发展的一个不容忽视的因素之一, 主要涉及到带宽与间断连接等问题.

3.2 数据挖掘技术

大数据的 4V 特性以及普适云的技术特征, 使得大数据时代的数据挖掘技术面临着新的挑战, 其所面临的主要挑战如下所述:

(1) 数据挖掘算法的改进. 现有的挖掘算法需要基于云计算进行改进. 目前, 基于云计算的并行数据挖掘系统研究, 大部分是在 Hadoop 的框架下展开的. 它是一个开源的可运行于大型分布式集群上的并行编程框架, 提供了一个支持 MapReduce 并行编程模型的部件. Hadoop 具有良好的存储和计算可扩展性; 具有分布式处理的可靠性和高效性; 具有良好的经济性(运行在普通 PC 机上). Hadoop 当前已经发展为一个项目集合, 最核心的是存储模块 HDFS 和计算模块 Map-Reduce, 此外还有 HBase、Hive、P ig、ZooKeeper、Cascading 等模块计算领域面临的存储 TB /PB 级数据, 以及对如此大的数据量进行高效、可靠、可扩展计算的诸多问题. 一些研究者在 Hadoop 架构下提出了 PFP 算法, MPFP 算法、BPFP 算法、基于云计算的 SPRINT 算法与基于云计算的 K-MEANS 算法, 等等. 此外, APS(Apache Software Foundation) 旗下的开源项目 Mahout, 基于 Apache Hadoop 库, 也可以实现大规模数据上的并行数据挖掘, 包括频繁模式挖掘、聚类与分类等算法, 且截至目前已发布多个版本.

(2) 数据类型的多样性. 不同的终端, 由于生产标准的差异性, 产生不同结构的数据, 其中包括: 结构化数据, 半结构化数据和非结构化数据, 这些异构化数据的抽取与集成也将成为一大挑战.

(3) 数据噪声太大. 普适终端的所处地理位置的复杂性, 使得数据具有很多噪声. 在进行数据清洗时, 不易把握清洗粒度. 粒度太大, 残留的噪声会干扰有价值的信息; 粒度太小, 可能会遗失有价值的信息.

(4) 数据的安全性与隐私保护[13]. 普适计算网络一般是以多种无线网和移动网接入互联网实现异构集成的网络. 无线网和移动网注定了其较低的可用性和可靠性, 而互联网的交互性, 使得人们在不同地点产

生的数据足迹得到积累和关联, 从而增加了隐私暴露的概率, 且这种隐性的数据暴露往往是无法控制和预知的. 从技术层面来说, 获取用户隐私的过程就是知识发现的过程. 所谓的“人肉搜索”即可类比成一个人力实施的大数据挖掘过程. Dwork 在 2006 年提出了新的差分隐私方法, 但这项技术离实际应用还很远.

(5) 挖掘结果的可视化. 可视化是最佳的结果展示方式之一, 但超大规模数据的可视化却面临着极大的挑战[14].

(6) 大数据挖掘的盲目性. 大数据挖掘不同于传统的数据挖掘, 其“先有数据后有模式”的特性, 使得大数据挖掘如同大海捞针, 盲目而复杂.

4 结束语

大数据挖掘研究, 不仅具有深远的科学研究价值, 而且可以产生巨大的经济效益和社会价值. 普适计算和云计算二者融合(简称为普适云 Ubicloud)为大数据挖掘提供了技术支撑与实现平台. 基于此, 本文提出了基于普适云的大数据挖掘架构, 在理论方面论证了其可行性,论述了其运行模式并进行了性能分析, 最后分析总结了基于普适云的大数据挖掘所涉及的关键技术. 下一步工作将集中在对该架构的实验验证, 通过实验分析其实用性与适用性.

参考文献

1 孟小峰.Cloud Computing and Big Data.北京:数据挖掘教学研讨会,2012.

2 孟小峰,慈祥.大数据管理:概念、技术与挑战.北京:中国人民大学信息学院,2012.

3 UN Global Pulse. Big Data for Development: Challenges & Opportunities.http://www.unglobalpulse.org/projects/BigDataforDevelopment.

4 陈援非,崔丽,朱珍民,曾益,吴元昆.UbiCloud:一种面向普适终端的云计算系统.计算机科学,2011,38(10):127–132.

5 亓开元,赵卓峰,房俊.基于 MapReduce 模型的大规模数据流处理.计算机学报,2012,35(3):476–489.

6 Weiss A. Computing in clouds. ACM Networker, 2007, 11(4): 18–25.

7 Buyya R, Yeo CS, Venugopal S. Market-oriented cloud computing: vision, hype, and reality for delivering IT services as computing utilities. Proc. of the 2008 10th IEEE International Conference on High Performance Computing and Communications. 2008, 0: 5–13.

8 Henricksen K, Indulska J. Infrastructure for pervasive computing: challenges. Proc.of the Informatik, Workshop on Pervasive Computing. 2001. 214–222.

9 焦扬.普适计算的现状及发展.http://wenku.baidu.com/view/bea52981bceb19e8b8f6ba7f.html.

10 Huifeng S, Yan L, Feng W. A high-performance remote computing platform. Pervasive Computing and Communications, 2009, PerCom 2009. IEEE International Conference on. 2009. 1–6.

11 牟权,叶保留,陆桑璐.基于云计算的普适服务集成平台技术研究.http://www.wenkudaquan.com/doc/20120603/260642.html.

12 童晓渝,张云勇,徐雷.智能普适网络架构及关键技术研究.http://blog.csdn.net/xiaogugood/article/details/7960717.

13 Lahlou S, Langheinrich M, Rocker C. Privacy and trust issues with invisible computers. Communications of the ACM, 2005, 48(3): 59–60.

14 Wong PC, Shen HW, Johnson CR. The top 10 challenges in extreme scale visual analytics. Computer Graphics and Applications, 2012, 32(4): 63–67.