

泛在网络空间大数据“雾云计算”软件体系结构

贾焰¹, 方滨兴², 汪祥¹, 王永恒³, 安静斌³, 李爱平¹, 周斌¹

(1. 国防科技大学, 长沙 410073; 2. 中国工程院, 北京 100088; 3. 之江实验室, 杭州 310012)

摘要: 随着网络空间从传统互联网向人、机、物、服务与应用互联的泛在网络空间扩展, 计算模式从“以云端集中计算为中心”向“前端、中间层和云端相结合”的方式转变, 传统的云计算、边缘计算等计算模式难以满足泛在网络空间大数据计算需求。本文针对泛在网络空间大数据计算面临的问题, 在知识体基础上, 提出了一种泛在网络空间大数据“雾云计算”软件体系结构, 在多知识体协同计算语言与模型基础上, 实现雾端、中间层和云端多知识体的协同计算, 为泛在网络空间大数据计算提供解决方案。

关键词: 雾云计算; 泛在网络空间; 大数据; 物联网; 云计算

中图分类号: TP319 **文献标识码:** A

Software Architecture of Fogcloud Computing for Big Data in Ubiquitous Cyberspace

Jia Yan¹, Fang Binxing², Wang Xiang¹, Wang Yongheng³, An Jingbin³,
Li Aiping¹, Zhou Bin¹

(1. National University of Defense Technology, Changsha 410073, China; 2. Chinese Academy of Engineering, Beijing 100088, China; 3. Zhejiang Lab, Hangzhou 310012, China)

Abstract: The cyberspace has expanded from traditional internet to ubiquitous cyberspace which interconnects human, machines, things, services, and applications. The computing paradigm is also shifting from centralized computing in the cloud to combined computing in the front end, middle layer, and cloud. Therefore, traditional computing paradigms such as cloud computing and edge computing can no longer satisfy the evolving computing needs of big data in ubiquitous cyberspace. This paper presents a computing architecture named Fogcloud Computing for big data in ubiquitous cyberspace. Collaborative computing by multiple knowledge actors in the fog, middle layer, and cloud is realized based on the collaborative computing language and models, thereby providing a solution for big data computing in ubiquitous cyberspace.

Keywords: fogcloud computing; ubiquitous cyberspace; big data; Internet of Things; cloud computing

收稿日期: 2018-08-31; 修回日期: 2019-08-15

通讯作者: 汪祥, 国防科技大学助理研究员, 主要研究方向为数据挖掘; E-mail: xiangwangcn@nudt.edu.cn

资助项目: 中国工程院咨询项目(2016-ZCQ-10); 自然科学基金重点项目“面向复杂查询的异质媒体搜索理论与方法研究”(61732004); 自然科学基金面上项目“面向云计算平台的数据安全与隐私保护关键技术研究”(61472433); 自然科学基金重点项目“泛在网络安全搜索基础理论与技术”(61732022)

本刊网址: www.engineering.org.cn/ch/journal/sscae

一、前言

信息技术是推动各行业发展的催化剂,随着网络空间从面向信息的互联网扩展到了人、机、物、服务与应用互联的泛在网络空间,传统的高性能计算 [1]、云计算 [2]、雾计算 [3]、边缘计算 [4] 和多 Agent 系统 [5,6] 等计算模式已经不能满足网络空间大数据分析处理的需要。

(一) 网络空间的扩展

互联网、移动互联网、物联网和社交网络等的迅猛发展使得网络空间日趋繁荣,促进网络空间由单一互联网发展到了泛在网络空间。泛在网络空间是一种建立在互联网基础之上的具有自适应性的智能网络,通过各种有线和无线网络与物联网、互联网、传感网等进行融合,综合利用海量的传感器、智能处理设备等终端以及运行在这些终端上面的软件、服务和应用,实现人、机、物、信息在任何时间、任何地点的安全有效联接。网络空间的扩展,使得传统“以后台处理为中心”的云计算模式不能满足“互联网+物联网+移动互联网”所形成的泛在网络空间计算需求,计算模式需要从“以后台处理为中心”向“后台与边缘、中间相结合”的方式转变,这是必然发生的革命性变革。

(二) 大数据特征的要求

大数据 [7] 是指在一定时间范围内难以依靠已有数据处理技术进行有效采集、管理和分析的数据集合,以下从其 5V 特征分析网络空间大数据面临的挑战:

(1) 规模大 (Volume)。从太字节 (TB) 级跃升到拍字节 (PB) 级甚至皆字节 (ZB) 级,单一的应用程序和数据中心难以快速处理如此海量的数据,迫切需要像人类一样由多知识体协同处理。

(2) 数据类型繁多 (Variety)。如文本、视频、音频、图片、特定类型等,甚至包括这些文件类型的交叉融合。单一的个人、公司、研究机构等难以同时掌握不同类型数据的分析处理工作,迫切需要创建一种协同合作并保护各自知识产权和数据所有权的计算模式。

(3) 速度快 (Velocity)。数据高速生成,要求实时处理,迫切需要在数据的源头和离源头不远的

边缘端或中间层进行数据处理,并与云端保持知识和信息交互。

(4) 数据的准确性和可信性 (Veracity)。数据及其来源存在不确定和不可信,迫切需要对不同来源的数据进行可信计算与交叉验证。

(5) 有价值 (Value)。大量的数据中存在极有价值的信息,而价值的挖掘通常与领域相关,需要各行各业的领域专家相互配合协同挖掘大数据中的价值。

因此,泛在网络空间大数据迫切需要大量个人、研究机构和企业等相互合作,在保护各自知识产权、数据隐私与所有权的基础上,进行综合的协同计算,这是当前计算模式难以满足的。

(三) “信息孤岛”与产权保护

网络空间大数据尤其是物联网数据、传感器数据和移动应用数据等大多为不同的个人、组织或机构所有。由于隐私保护、所有权保护和保密等原因,当前大数据“信息孤岛”现象明显,数据和知识难以集中到某个数据中心进行统一的存储、搜索、分析和挖掘 [8]。在数据分析处理挖掘方面,分散在全世界各个研究机构和组织的专家在各自的行业和研究领域取得了丰硕的成果,但难以出现一家单位和机构能够独自解决当前泛在网络空间大数据分析处理面临的技术难题。迫切需要在保护各自知识产权的基础上进行合作,使得世界上的优秀研究成果能够协同计算,满足泛在网络空间大数据分析处理的需要。

为了解决上述问题,本文提出一种面向泛在网络空间大数据的通用智能化“雾云计算”体系结构,通过前端(雾端)、中间层和云端知识体以及它们之间的协同计算,满足网络空间大数据在产权保护基础上的实时及融合分析处理需求。本文知识体是一个自治的、智慧的软件实体,其逻辑上有自己的领域、目标和功能,物理上有自己的知识库、学习器、推理器、执行器和协同器。知识体不仅可以根据目标构建知识,并根据任务需要提供功能,还可以互相协同计算,共同完成某个任务。雾云计算体系结构可以应用在物联网、智能家居、问答机器人等泛在网络大数据计算场景,其中网络空间大搜索 [9] 是雾云计算应用的典型案例。当前,苹果移动操作系统 (iOS) 中的智能语音助手 (Siri) [10]、

华为语音助手 HiVoice [11] 等产品将本地手机中的应用软件 (APP)、用户个人隐私数据与云端互联网数据等进行协同计算, 满足用户智慧解答的需求, 体现了泛在网络空间“雾云计算”的基本思想。

二、雾云计算基本概念

雾云计算通过对大数据海洋中抽取知识的雾端、知识融合的中间层和智慧计算的云端多知识体进行协同计算, 获得满足用户计算目标的过程。雾云计算由知识体、知识体之间的关系和知识体之间的协同操作 3 部分组成, 如式 1 所示:

$$\text{FogcloudComputing} = \{\text{KnowledgeActor}, R, \text{Operation}\} \quad (1)$$

式 (1) 中, KnowledgeActor 表示知识体的集合, 单个知识体可用 KnowledgeActor_{*i*} (*i* = 1, ..., *n*) 表示; *R* 表示知识体之间的关系集合, *R*_{*jk*} 表示知识体 KnowledgeActor_{*i*} 和 KnowledgeActor_{*k*} 之间的关系, *j, k* = 1, ..., *n*。Operation 表示知识体之间的协同操作与计算, Operation_{*m*} 表示与多知识体协同工作的任务 Task_{*m*} 相关的协同操作与计算。

知识体是一个智慧的软构件, 可以是雾端面向目标的简单知识抽取软构件, 也可以是云端的超大规模知识体, 如健康知识体、天气预报知识体、人类常识知识体等。知识体具有自学习、自演化、可描述、可管理、灵活在线组装、分布式部署等特点, 并且知识体之间可以进行协调问题求解。某个知识体 KnowledgeActor_{*i*} = {*ID*, *O*, *A*} 由标识 *ID*、目标 Object 和内部架构 Architecture 三部分组成, 其中,

目标 *O* 可表示为:

$$O = \{\text{CollectData}, \text{KnowledgeAcquisition}, \text{Inference}, \text{Computing}, \text{Merging}, \dots\} \quad (2)$$

式 (2) 中, 目标 *O* 包括数据获取 (Collect Data)、知识获取 (Knowledge Acquisition)、知识推理 (Inference)、知识计算 (Computing) 和知识融合 (Merging) 等。

知识体的体系结构如图 1 所示, 包括任务执行器 (Task Executor)、知识库 (Knowledge Base)、推理器 (InferEngine)、学习器 (Learn Unit)、任务规划及决策器 (TPDK)、协同器 (Cooperator)、知识输入输出接口 (KInterface)、协同接口 (CInterface)、数据接口 (DInterface)、任务输入接口 (TInterface) 等。知识体的体系结构 *A* 可以用下面公式形式化表示:

$$A = \{\text{TaskExecutor}, \text{KnowledgeBase}, \text{InferEngine}, \text{LearnUnit}, \text{TPDK}, \text{CInterface}, \text{DInterface}, \text{TInterface}\} \quad (3)$$

三、雾云计算体系结构

泛在网络空间大数据“雾云计算”体系结构如图 2 所示, 在多知识体及其协同推理的基础上, 由雾端的前端计算、中间层的融合计算和后端的云计算组成, 其中各层知识体分别运行在不同的知识体运行平台上。与边缘计算、云计算的主要区别是: 在多知识体的基础上, 支持前端 (雾端)、中间和后端 (云端) 的高度协同计算能力。

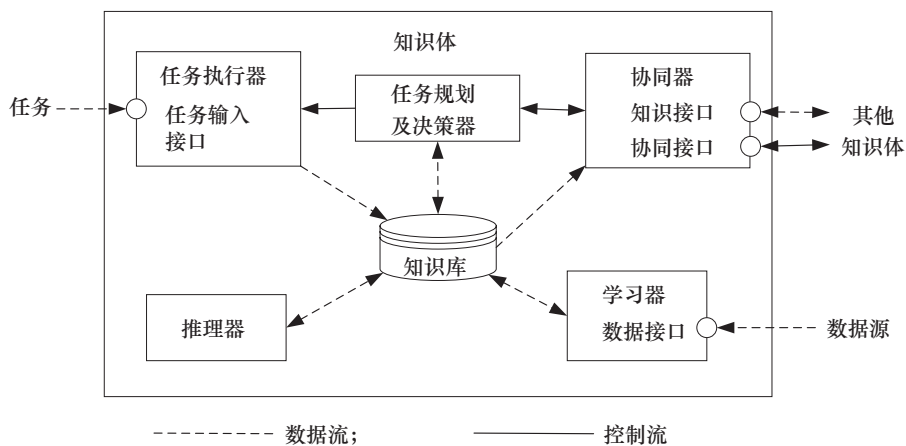


图 1 知识体系统结构

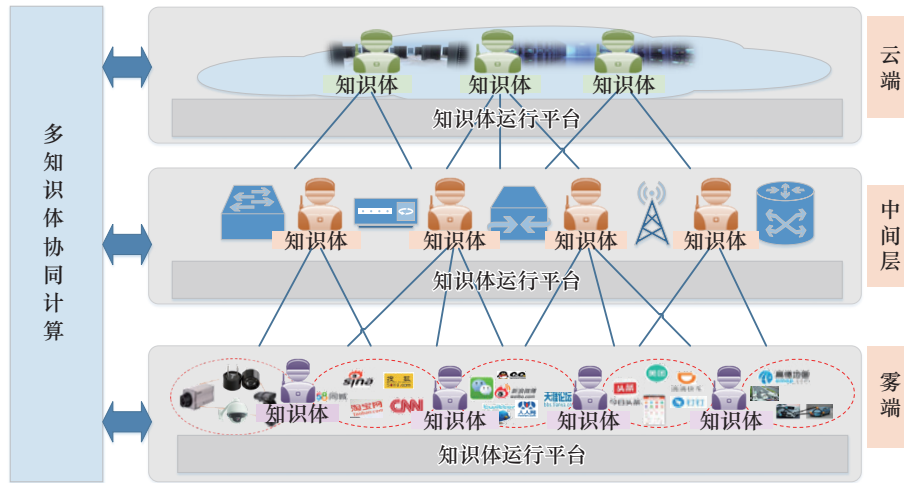


图2 雾云计算体系结构

（一）云端知识体

云端知识体通常是雾云计算的远端后台处理中心，是分散在世界范围内的边缘物联网节点、网页（Web）应用、移动APP、互联网中间节点知识的汇集地，也是分散在海量雾端和中间层知识体内容的集散地。云端知识体对雾端与中间层知识体上下文和时空位置敏感，是雾云计算的全局中心。云端知识体通常存储非隐私的全局和上下文数据，并进行知识的全局推理。

云端知识体的特点主要表现在以下几个方面：

①计算和存储能力非常强，网络非常稳定；②通常是对分散的中间层知识体和雾端知识体数据进行全局性的综合分析计算；③获得的知识分布范围广，通常需要进行较大空间尺度的时空分析计算；④需要与中间节点知识体和雾端节点知识体进行协同推理计算；⑤由于数据量庞大和计算复杂，通常进行离线分析计算。

云端知识体通常运行在以下计算集群与平台基础上。

1. 分布存储集群

雾云计算平台面向泛在网络空间大数据计算任务，从遍布互联网、物联网的海量数据源获取信息并经过各种功能知识体抽取、加工、融合，形成知识，并在此基础上根据具体任务进行计算。作为运行大量各类具备知识和智能知识体的运行平台，必然需要存储大量结构化、半结构化和非结构化的数据，包括文件、网页、知识图谱、音视频等。分布存储集群由若干存储服务器和索引服务器组成，采用分布式文件系统和图数据库等专用数据存储管理

系统构建。

2. 人工智能（AI）计算集群

AI计算集群用于云端知识体训练图片、音视频等高密度计算任务。AI计算集群采用中央处理器+图形处理器（CPU+GPU）的架构，由若干高密度计算服务器节点通过高速互联构成。

3. 流数据处理集群

流数据处理集群用于对来自前端节点的数据源和知识体获取的海量数据流进行实时处理，这些数据流可能来自于传感器、移动终端，也可能来自于远程部署的智能爬虫。海量的流数据需要实时进行清洗、筛选和计算，然后存储到分布存储集群作为知识发现的素材，或者直接展现给搜索请求用户知识发现。

4. 高速缓存集群

高速缓存集群用于存储索引、热点数据等需要快速访问的数据，主要采用内存存储，为高实时性计算和需要快速响应的用户请求提供支持。

5. 消息交换集群

消息交换集群为大量知识体间协同计算提供底层通信基础设施，集群服务器节点配置高速的网络接口带宽。

（二）中间层知识体

中间层知识体位于雾端和云端之间，通常部署在离边缘节点较近的服务器上，有时也会部署在网关、路由器、通信基站等汇聚型节点上。中间层知识体通常运行在计算能力、存储能力较强的服务器上，相对于云端节点来说，更贴近雾端边缘节点和

数据源，是局部边缘知识体综合计算的中心。中间层知识体在雾云计算中扮演核心角色：一方面，中间层知识体能够访问互联网中的海量数据，并与云端知识体进行协同计算；另一方面，中间层知识体能够访问多个本地边缘节点中的大量数据和知识（这些数据和知识通常是难以快速返回云端中心的信息，例如用户的隐私信息通常希望存储在本地、实时的视频监控数据由于数据量过大而难以实时传输到云端计算中心），可以基于边缘节点地理位置的分布进行协同计算。**中间层知识体是全局网络空间计算和本地数据结合的重要区域，是雾云计算的核心。**例如，在一个家庭的雾云计算中间节点，存储着大量的家庭传感器数据、视频监控数据和移动应用数据等，为了提供实时的综合反馈和保护家庭隐私，本地的雾云计算中间节点将对这些数据与云端节点数据进行融合计算和实时反馈。

中间层知识体的特点主要表现在以下几个方面：①计算和存储能力较强，能够进行较复杂的综合计算分析，网络较稳定；②贴近数据源，能够基于地理位置的特点进行时空分析计算和融合；③实现多个边缘节点数据（如视频监控等不需要存储在云端的数据、隐私数据等）和云端知识体的协同推理计算；④连接的雾端边缘节点数据和知识具有异质性特点，对融合分析计算提出了严峻挑战；⑤通常进行近实时的分析计算和反馈。

（三）雾端知识体

数量巨大的雾端边缘节点是数据的直接来源，包括智能物联网中的设备、传感器网络中的传感器、移动互联网中的APP、智能家居设备、安装智能设备的交通工具（汽车、无人机等）等。这些边缘节点中有很多也具备一定的计算能力，可用于部署本地实时计算和数据及知识获取知识体，是泛在网络空间大数据计算的一部分。

雾端知识体的特点主要表现为：①计算和存储能力较弱，网络不够稳定；②是数据的直接来源；③具有时间和空间位置信息；④能够在较大范围内进行移动；⑤知识体通常进行实时计算和反馈。

（四）多知识体协同计算

分布在雾端、中间层、云端的多知识体，在协同计算语言的支持下协同计算与任务调度。

多知识体协同计算语言对领域（Domain）、知识体（KnowledgeActor）、行动（Action）、问题（Problem）、目标（Goal）等一系列对象进行定义。多知识体协同任务规划由多目标优化的任务规划器生成，在知识体创建和组装后，基于多知识体协同任务规划生成多知识体的具体协同策略。针对多知识体协同计算任务，多目标优化器从任务规划空间求解优化的任务规划，根据生成的任务规划，由知识体组装及部署器进行知识体的在线组装，并按需部署到计算节点。多知识体的协同既有层层间的协同（如从雾端向中间层和云端汇聚知识），也有同层之间的协同（如对不同数据和知识的计算进行分工合作）。多知识体的协同也是一种动态的协同，在任务执行期间，根据系统的负载情况或者知识源的变化情况，知识体可以动态地加入或退出任务。

四、结语

本文针对泛在网络空间大数据应用需要在云端、中间层和前端进行计算的问题，提出了一种基于知识体的云端、中间层和前端协同计算的“雾云计算”软件体系结构。描述了“雾云计算”的基本概念，阐明了“雾云计算”体系结构的基本组成及其关系，其关键技术包括多层次多粒度多模态数据表示与知识推理、多知识体协同推理、时空融合分析计算、可信搜索与隐私保护等。

“雾云计算”体系结构适用于典型的泛在网络空间大数据计算应用，为泛在网络空间中前端、中间层和云端的协同计算提供了解决方案。“雾云计算”可应用在物联网、智能家居、问答机器人等泛在网络大数据计算场景，其中网络空间大搜索是雾云计算的典型用例。网络空间大搜索面向泛在网络空间中的人、物体、信息和服务，在正确理解用户意图的基础上，基于从泛在网络空间大数据获取的知识，给出满足用户需求的智慧解答。

参考文献

- [1] Schroeder B, Gibson G. A large-scale study of failures in high-performance computing systems [J]. IEEE Transactions on Dependable and Secure Computing, 2009, 7(4): 337-350.
- [2] 罗军舟, 金嘉晖, 宋爱波, 等. 云计算: 体系架构与关键技术 [J]. 通信学报, 2011 (7): 3-21.
Luo J Z, Jin J H, Song A B, et al. Cloud computing: Architecture

- and key technologies [J]. *Journal on Communications*, 2011 (7): 3–21.
- [3] Bonomi F, Milito R, Zhu J, et al. Fog computing and its role in the internet of things [C]. Helsinki: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, 2012.
- [4] Shi W, Cao J, Zhang Q, et al. Edge computing: Vision and challenges [J]. *IEEE Internet of Things Journal*, 2016, 3(5): 637–646.
- [5] 刘大有, 杨鲲, 陈建中. Agent研究现状与发展趋势 [J]. *软件学报*, 2000, 11(3): 315–321.
Liu D Y, Yang K, Chen J Z. Agents: Present status and trends [J]. *Journal of Software*, 2000, 11(3): 315–321.
- [6] Olfati-Saber R, Fax J A, Murray R M. Consensus and cooperation in networked multi-agent systems [J]. *Proceedings of the IEEE*, 2007, 95(1): 215–233.
- [7] Mayer-Schonberger V, Cukier K. Big data: A revolution that will transform how we live, work, and think [M]. Boston: Houghton Mifflin Harcourt, 2013.
- [8] Sicari S, Rizzardi A, Grieco L A, et al. Security, privacy and trust in Internet of Things: The road ahead [J]. *Computer Networks*, 2015, 76(15): 146–164.
- [9] Fang B X, Jia Y, Li X, et al. Big search in cyberspace [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(9): 1793–1805.
- [10] López G, Quesada L, Guerrero L A. Alexa vs. Siri vs. Cortana vs. Google Assistant: A comparison of speech-based natural user interfaces [C]. Los Angeles: International Conference on Applied Human Factors and Ergonomics, 2017.
- [11] Baker M. Real, unreal, and hacked [J]. *IEEE Pervasive Computing*, 2018, 17(1): 104–112.