基于海云协同的物联网大数据管理

赵永波[1,3] 陈曙东[2,3] 管江华[1,3] 褚 震[3] 杨 萃[4]

[1](中国科学院大学 北京 100049)
[2](中国科学院微电子研究所 北京 100029)
[3](中国物联网研究发展中心 无锡 214135)
[4](中国科学院计算机网络信息中心 北京 100190)

摘 要 大数据不断地从复杂的应用系统中产生，并且将会以更多、更复杂、更多样化的方式持续增长。多样化的物联网传感设备不断地感知着海量的具有不同格式的数据。物联网系统中大数据的复杂化和格式多样化，决定了物联网系统中针对大数据的应用场景和服务类型的多样化，从而要求物联网大数据管理系统必须采用不同的新技术来应对具有不同格式的大数据，而现有的针对特定数据类型和业务的系统在架构上已经难以满足如此多样化的需求，因此，设计新的具有可扩展性的系统架构已经成为物联网大数据管理的研究热点。文章提出了一种物联网大数据管理的创新解决方案：面向物联网大数据管理的海云协同模型。首先讨论海云协同模型的整体架构和协同机制，然后分别讨论了海云协同模型中海端计算系统和云端计算系统的设计和实现方案，测试结果表明提出的解决方案性能良好、具有实践可行性。

关键词 海云协同；物联网；大数据；大数据管理
中图分类号 TG 156 **文献标志码** A

Managing Big Data of Internet of Things Based on Sea-Cloud Synergy

ZHAO Yongbo[1,3] CHEN Shudong[2,3] GUAN Jianghua[1,3] CHU Zhen[3] YANG Cui[4]

[1](*University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)
[2](*Institute of Microelectronics*, *Chinese Academy of Sciences*, *Beijing* 100029, *China*)
[3](*China R&D Center for Internet of Things*, *Wuxi* 214135, *China*)
[4](*Computer Network Information Center*, *Chinese Academy of Sciences*, *Beijing* 510006, *China*)

Abstract Big data is generated from sophisticated applications and is going to continue growing over the future years in a much diverse, larger and faster manner. Particularly, massive data with different formats sensed by various sensors, devices were from independent or connected components of Internet of Things (IoT) applications decide the diversity of applications and service type in IoT systems, which requires new technologies to manage big data with various format and to meet the requirements of various IoT applications. For such diverse business requirements, traditional systems are difficult to give a preferable solution due to their incapable system architectures. Consequently, to design new scalable system architectures is becoming a key research point for IoT big data management. This paper aims to provide a novel solution to manage big data of IoT systems, namely, the sea-cloud synergy model. Firstly, the designed architecture and the synergy mechanism of the model were discussed. Secondly, the design and implementation details of sea-side and cloud-side of the sea-cloud synergy model were presented respectively. Finally, a demonstration system was built . Experimental experimental results verifies the preferable performance and feasibility of the designed sea-cloud synergy model.

Keywords sea-cloud synergy; internet of things; big data; data management

收稿日期：2014-03-04

作者简介：赵永波，硕士研究生，研究方向为大数据管理与数据挖掘；陈曙东(通讯作者)，博士，研究员，博士研究生导师，研究方向为大数据管理，E-mail：chenshudong@ciotc.org；管江华，硕士研究生，研究方向为大数据分析与云结算；褚震，工程师，研究方向为云计算与分布式系统；杨萃，硕士研究生，研究方向为物联网。

1 引 言

近年来，由各种各样传感设备所感知，且具有不同格式的数据迅猛增长。多样化的社交应用，如 Facebook 和 Twitter 等所产生的大量用户数据也在不断加速这种增长。物联网系统[1]尤其是移动物联网系统由数十亿的无线传感组件构成，这些组件时刻执行着感知、收集和处理具有不同类型数据的任务，可以预见，随着时间的推移，这些物联网应用将推动数据空间达到更大的规模。在物联网系统中，人和设备(从智能手机到可穿戴智能设备，从安装在汽车里的智能传感器到宇宙飞船)紧密互联，从这数十亿互联的组件中产生的大量传感数据将产生一个巨大的数据海洋，从而加速了大数据[2, 3]的出现。从如此大量的数据中提取有价值的信息来提高日常生活质量和办公效率是信息科技发展的必然需求。

以交通问题为例，假设你在环山公路上开车，在进入某个拐角的指定区域时，恰好有一辆车位于拐角另一端被困在交通事故中，这辆车将会给你的车发送一条事故预警信息以提醒你减速慢行。这种基于动态自组织网络技术[4, 5]的碰撞预警系统将会在一瞬间保护我们远离撞车甚至是伤亡。另一个例子，一个位于云端的智能交通系统[6]将会实时地计算出一条从家到办公室的最近的路线，同时这条路线具有尽可能少的交通拥堵，然后云端系统会将这条路线推送到用户智能手机上的客户端应用程序。此外，智能交通系统将会区分那些要求实时计算和响应的任务(如碰撞预警系统)以及那些依赖大量数据资源和复杂计算过程的任务(如天气预报和交通流量预测)，然后将这些具有不同需求的任务指派到不同的端系统执行(如本地计算节点和云端系统)。在所有这些物联网系统应用场景中，对于利用大数据解决实际问题这一现实需求，我们都面临着以下这几方面的挑战：

(1)需要新的系统架构来管理整个信息空间的大数据的生命周期。

(2)需要新的协同机制来判断具有不同实际需求的任务，然后将这些任务指派到相应的端系统执行。

(3)需要新的存储和计算技术来完成对大数据的存储和分析工作，从而进一步地挖掘大数据的潜在价值来生成按需获取的服务信息，同时确保服务信息的实时响应。

鉴于以上挑战的存在，本文提出一种创新的物联网大数据管理解决方案，即面向物联网大数据管理的海云协同模型。

文章结构如下：第 2 部分讨论海云协同模型的整体架构和协同机制；第 3 部分讨论海云协同模型中海端计算系统的设计方案；第 4 部分讨论云端物联网大数据管理系统的设计和实现方案，即海云协同模型中云端系统的实现；第 5 部分介绍测试案例并对结果进行研究分析；相关的研究工作和总结在第 6 和第 7 部分给出。

2 海云协同模型

物联网系统具有显著的异构性、混杂性和超大规模等特性。异构性表现在不同的制造商、不同的拥有者、不同的类型以及不同范畴的对象网络共存于物联网中；混杂性表现在网络形态、组成、场景、服务和应用等多个方面；超大规模性表现在物联网系统是物理世界与信息空间的深度融合，是全球范围的人、机、物的互联。所有这些物联网系统的特性决定了物联网传感数据也具有异构性、混杂性、实时感知和超大规模等特性。这些特性决定了众多不同的物联网大数据应用场景中大数据处理任务的异构性和需求多样性，这些任务的异构性和需求的多样性要求物联网大数据管理系统必须采用不同的新技术来处理具有不同格式的大数

据，而现有的针对特定数据类型和业务的系统在架构上已经难以满足如此多样化的需求。这意味着我们需要设计新的系统架构，不仅要满足实时响应服务但依赖较少数据和计算资源任务的需求，而且能满足依赖大量数据和计算资源但不要求实时响应服务的需求。

基于以上实际需求，本文提出了面向物联网大数据管理的海云协同模型，图 1 为模型的整体架构。

如图 1 所示，海云协同模型的核心为定义了三种不同的服务请求类型，即：海端实时响应服务请求、云端实时响应服务请求和云端大数据分析与挖掘服务请求。

2.1　海端实时响应服务请求

如图 1 所示，海端实时响应服务判定器用于判断服务请求是否属于要求强实时服务响应的请求类型，并且这种请求只需依赖较少的数据量和计算量，但却要求很强的响应实时性。本文所讨论的海云协同模型中，响应这种服务请求的计算任务被定义为海端实时响应任务。例如车联网系统中车辆碰撞预警系统。在车辆碰撞预警系统中，当安装在车辆上的无线传感器感知到有其他车辆进入自身周围的特定范围内并且有可能发生碰撞时，会利用动态自组织网络技术与其他车辆

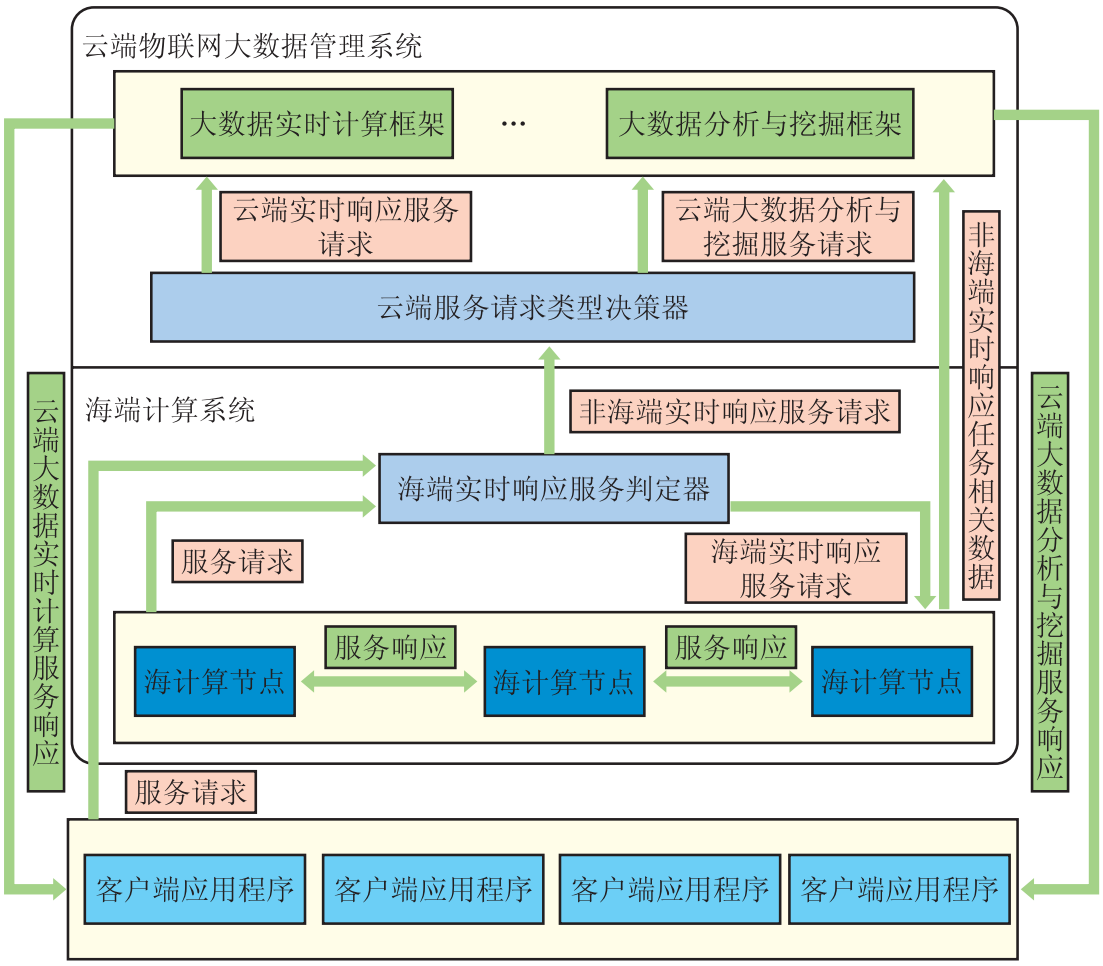

图 1　海云协同模型整体架构图

Fig. 1.　The architecture of the sea-cloud synergy model

组成一个动态临时网络，并且实时发送一条碰撞预警信息给其他车辆，以避免发生碰撞。碰撞危险解除后，传感器会将此次预警过程中的数据信息发送到云端物联网大数据管理系统，为道路整改或者交通流量控制等方案的设计提供有用的历史参考数据。这些方案的设计，可能需要经过大量的历史数据分析和挖掘以得出较好的方案，因此这种碰撞预警发生的时间、地点、周边交通状况等数据信息将会变得很有价值。自组网请求信息和碰撞预警信息只承载很少的数据量，并且可以在动态局部自组网中快速发送，同时海端实时响应服务判定器也会在这个局部动态自组网中实现，以降低与云端服务器交互而带来的网络负载，从而保证服务响应的高效实时性。如果服务请求属于这种类型，则完成服务响应的计算任务将直接在位于本地的物联网计算节点中执行，并实时给予其他本地计算节点或客户端应用程序以服务响应，这种响应是秒级甚至是毫秒级的。本文所讨论的海云协同模型中，这种本地物联网计算节点被称为海计算节点，而完成这种服务请求任务的计算过程被称为海计算，在车辆碰撞预警系统中，安装在车辆上的无线传感器就充当了海计算节点。

2.2 云端实时响应服务请求

如果服务请求不属于海端实时响应服务请求，则服务请求会被发送至云端物联网大数据管理系统，由云端服务请求类型决策器来判断服务请求的类型。此时会有两种判定的服务请求类型：云端实时响应服务请求和云端大数据分析与挖掘服务请求。云端实时响应服务请求是指那些既要求实时响应同时又依赖较多的数据和计算资源的服务请求，例如实时路径导航，实时交通拥堵状况查询等服务请求，这些应用请求可能来自用户智能手机上的客户端应用程序，实时路径导航需要依赖用户当前的位置和地图数据来实时计算出导航路线并且推送到客户端应用程序，这需要 GPS 定位数据和地图数据，并且经过一系列计算来完成导航。对于低功耗的海计算节点(通常为无线传感器节点)，其计算和存储能力显然不能负荷这样的计算任务，因此这类服务请求会被发送到云端计算系统进行实时计算(通常采用基于分布式系统的实时流计算技术)。

2.3 云端大数据分析与挖掘服务请求

云端大数据分析与挖掘服务请求是指那些依赖海量数据和计算资源才能完成响应的服务请求(通常为海量的数据和复杂的数学模型，如某种机器学习模型或者数据挖掘方法)。这些大数据的分析和挖掘工作是一个长期的工作，因此没有实时性的要求，从而可以在云端分布式计算系统中以离线的方式进行。在需要分析和挖掘结果时，利用大数据可视化工具呈现分析结果并且推送到客户端程序。例如利用出租车公司的历史载客数据，分析出租车乘客的区域密度分布以指导出租公司进行车辆分布规划；利用电子商务网站的用户历史购物数据分析用户的潜在购物兴趣，从而向用户推荐相关商品，促成潜在购物行为；利用大量的历史天气数据来预测未来的天气情况；利用历史路况信息和道路交通事故历史数据挖掘出有用的知识信息，给道路整改方案和交通流量控制方案的设计提供有价值的参考等。

2.4 海云协同模型的协同机制

海云协同模型协同机制的核心在于如何确定服务请求的类型，下面介绍海云系统模型协同机制的运行原理：

(1)在海云协同模型中定义了一组服务请求类型集，这组服务类型集是一组特定的服务请求类型和处理这个服务请求所需计算任务的映射集。

(2)当客户端服务请求被提交到服务请求类型决策器时(海端实时响应服务判定器和云端服务请求类型决策器)，决策器会解析出描述服务请求的类型标识符。

(3)服务请求类型决策器利用解析到的服务

请求类型标识符，搜索已定义的服务请求类型集来确定所需的计算任务。

(4) 服务请求类型决策器把所确定的计算任务提交到相应的计算模块(海计算节点、大数据实时计算框架、大数据分析与挖掘框架)执行。

(5) 计算模块在完成计算任务后把结果以服务的形式推送给服务请求客户端(海计算节点、客户端应用程序)，完成对服务请求的响应。

3　海端计算系统

在物联网系统中存在实时数据感知和处理任务以及实时响应型服务请求，将这种类型的数据处理和服务响应任务都提交到云端执行会消耗时间和网络带宽。一方面由于物联网系统中实时感应数据异构且庞大，另一方面是由于实时响应服务的高度实时性要求。例如车辆碰撞预警系统中

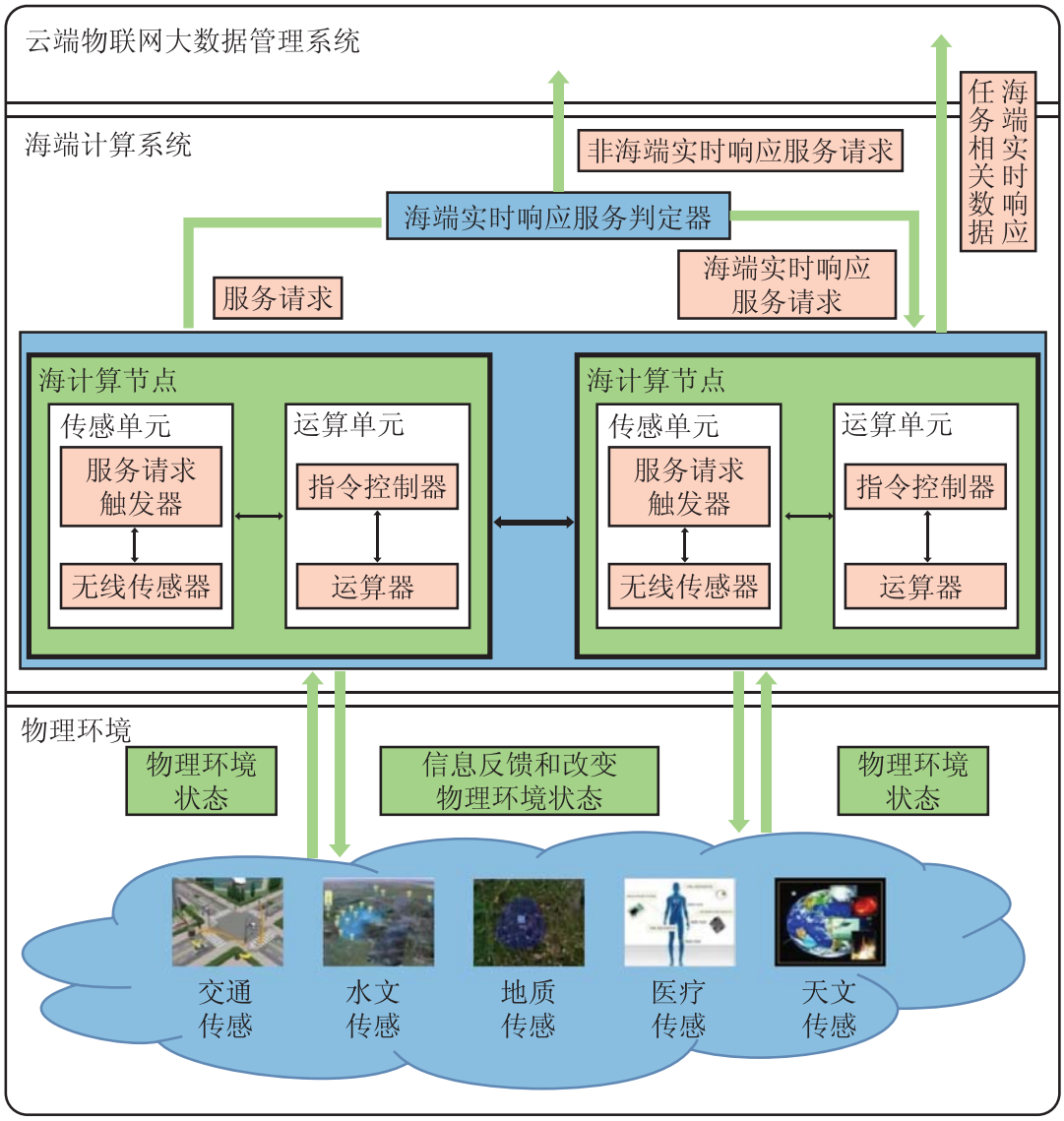

图 2　海端计算系统架构图

Fig. 2.　The architecture of the sea-side computing system

的实时服务响应。因此，海端计算系统的核心在于本地数据存储和计算，即在本地无线传感器节点中完成对实时感知数据的存储和处理工作，从而确保无线传感器节点之间的实时服务响应。例如：安装有无线传感器节点的一辆汽车正在靠近一个交通事故现场，此时服务响应信息为一条交通事故预警信息，提醒车辆提前减速慢行，绕开事故现场。海端计算系统的核心设计理念在于，在类似车辆碰撞预警系统的实际应用场景中，预警信息的产生并不需要依赖物联网系统的全局感知数据，而仅仅只是和事故现场车辆状态有关的局部传感信息，因此这些局部传感信息的处理完全可以独立于云端物联网大数据管理系统而在传感器自身完成。图 2 为海端计算系统的详细架构。

如图 2 所示，海计算节点由传感单元和运算单元构成，传感单元负责感知和存储物联网传感数据，而运算单元负责完成海端实时响应服务所需的数据处理和服务消息生成以及推送等任务，传感单元和运算单元协同完成海端动态自组织网络的构建以及海端实时服务响应等任务，即海计算。由于物联网应用的多样化，物联网系统将呈现一个可扩展的分布式的结构，因此海端计算系统不仅降低了网络流量的负载，同时海端计算系统的高度自治性将给物联网系统的扩展带来架构上的灵活性。必须强调，在海端计算系统中，由于海计算节点负责完成局部感知数据的存储和计算任务，因此设计具有更强存储和计算能力的无线传感器是海计算问题的重要部分。增强无线传感器的性能，属于电子设计的范畴[7]，不在本文的讨论范围之内。

4 云端物联网大数据管理系统

在海云协同模型中，海端计算系统针对那些依赖局部感知、本地存储和本地计算并且要求高实时性服务响应的物联网应用给出了解决方案。而云端物联网大数据管理系统的设计则是针对那些依赖全局感知、云端存储和海量复杂计算的物联网应用。在这些复杂的依赖海量数据和计算的应用中，海量感知数据有可能包括服务于交通流量预测的道路状况数据流；服务于病人状态监控和病情预测的医疗传感数据；服务于物流跟踪和客户兴趣分析的物流和商品零售感知数据流等。海量的复杂计算则可能包括利用大数据实时流计算技术实时计算出用户请求的最佳导航路径并推送给请求客户端；建立可行的机器学习模型利用大量的历史路况数据作为训练集来预测未来某个时间段的交通流量；利用 Apriori 算法[8]或 FP-Tree 算法[9]等数据挖掘方法从大量的电子商务网站交易记录中提取关联规则等。因此，设计高效的物联网大数据存取组件(如分布式存储系统 Amazon S3[10]、Google Bigtable[11]和 HDFS[12]等)，提出新的面向物联网大数据的实时计算技术，提出高效的面向物联网大数据分析和挖掘方法来满足上述实际应用的多样化需求，是设计云端物联网大数据管理系统的核心问题。

4.1 集成 OpenStack 和 Hadoop：云端物联网大数据管理系统解决方案

近年来，云计算受到了业界的高度关注并且成为分布式计算、资源共享、按需服务获取等问题的通用解决方案。OpenStack[13]是一个提供了基础设施即服务的 IaaS 软件项目，由控制了大量计算、存储和网络资源的一系列相互作用的组件构成。由于 HDFS 和 MapReduce 编程模型[14]的功能强大，Hadoop[15]已经成为大规模数据分析问题的首选工具。因此，我们使用 OpenStack 和 Hadoop 计算框架来构建我们的云端物联网大数据管理系统，图 3 为其系统架构。

如图 3 所示，云端系统实现了一系列旨在处理和响应云端服务请求(云端实时响应服务请求和云端大数据分析与挖掘服务请求)的服务组

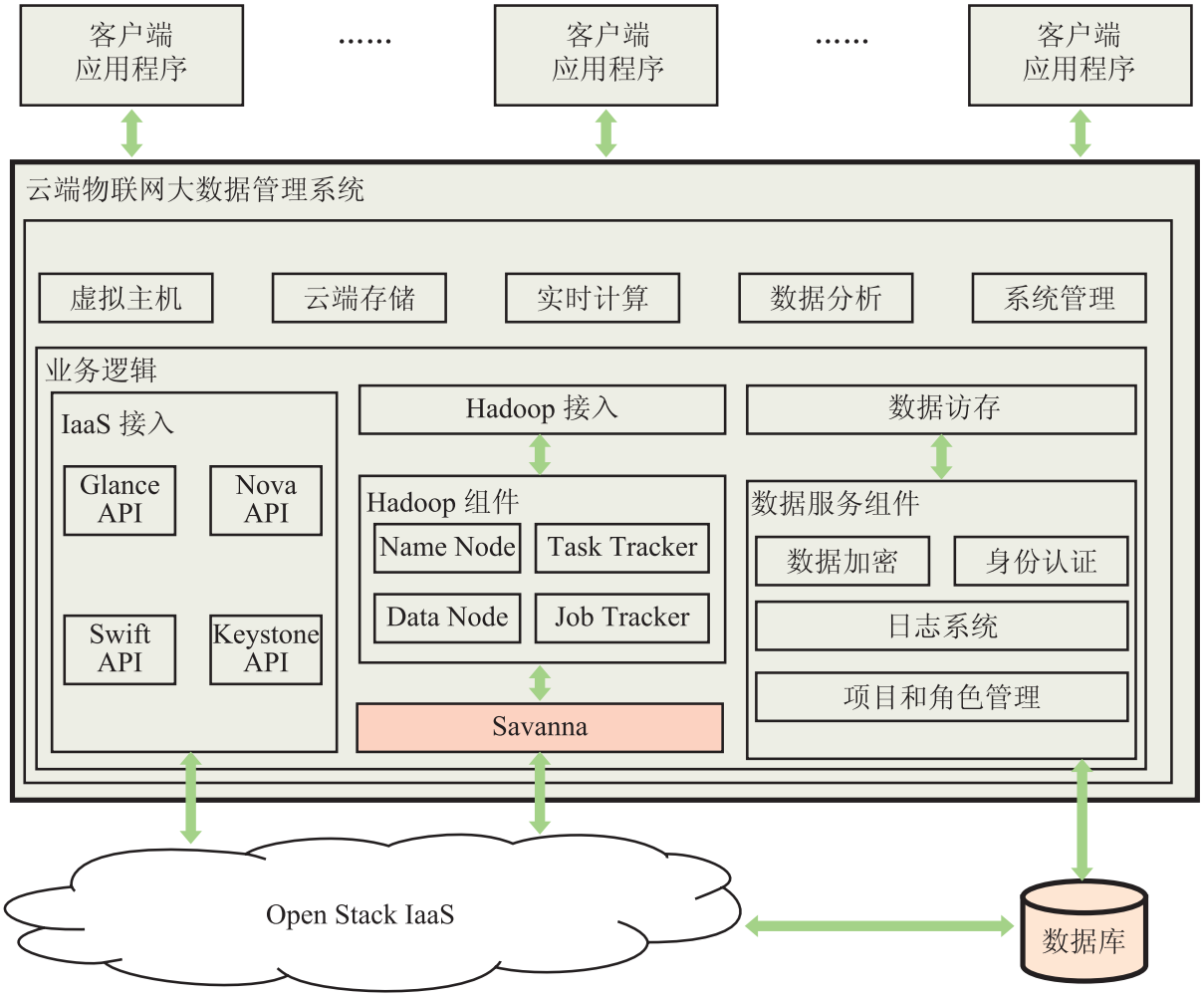

图 3　云端物联网大数据管理系统架构图

Fig. 3.　The architecture of cloud-side IoT big data management system

件。虚拟主机服务组件向终端用户共享了云端系统的硬件和软件资源，从而终端用户可以按需使用云端系统服务集群的计算和存储资源，完成数据和计算密集型大数据处理任务(天气预测等科学计算)；云存储服务组件提供大数据的快速存取服务，这些数据包括路况实时监测数据、交通事故相关数据等实时流数据和海量电子商务网站交易数据、历史天气数据和历史路况数据等，云存储服务组件由 OpenStack 的分布式对象存储系统 Swift 实现。实时计算服务组件执行云端实时服务响应请求所需的计算任务，包括实时路径导航服务请求等。数据分析服务组件使用一些列工具组件(数据集成和清理工具、大数据挖掘和机器学习算法库等)来执行云端大数据分析与挖掘服务请求所需的计算任务，实时计算服务组件和数据分析服务组件都由 Hadoop 计算框架实现。最后，系统管理服务组件执行与系统管理、服务租用和计费等业务相关的一系列服务。

由上述，云端系统实现的所有服务组件都是基于 OpenStack 和 Hadoop 搭建的，因此，在服务组件集之下是一组获取 OpenStack 和 Hadoop 核心服务的访问组件(即 Glance API、Nova

API、Swift API、KeyStone API 以及 Hadoop 接入组件)。数据访存模块和数据服务组件是一组实现用户认证、数据加密、系统服务计费等功能的系统支持组件。云端系统架构的最底层是 OpenStack IaaS 的核心组件以及用于用户认证、服务计费等服务的数据库系统。

基于此系统架构设计，我们开发了云端物联网大数据管理系统的原型系统——物联网大数据全生命周期管理系统，图 4 为系统的一个运行界面。

5 测试案例与分析

我们设计了两个测试案例来测试云端物联网大数据管理系统的实践可行性和性能表现，两个测试案例分别针对云端实时响应服务请求和云端大数据分析与挖掘服务请求的响应性能进行评估。

5.1 云端实时响应服务请求

测试采用实时获取道路拥堵状况对云端实时响应服务请求进行性能评估。

(1)测试环境搭建：采用的软件环境为云端物联网大数据管理系统中配置两个计算节点的 Hadoop 集群，硬件配置为配有 Intel Core i3 系列的双核 CPU(主频：3.30 GHz)以及 2 GB 内存的 PC 机(Hadoop 计算节点的硬件实现)。

(2)测试用例设计：用任一时间段内(本实验中为 3 分钟)某城市指定路段上所有车辆的平均速度来量化道路的拥堵状况，平均车速越低则表示道路拥堵越严重。在实验中，城市中所有车辆每隔 3 秒钟向云端系统上载自己的当前位置和速度数据，系统在收集到这些数据后，首先筛选出 3 分钟内某指定路段上的所有车辆的速度数据(依靠位置数据来筛选)，然后求出所有速度数据的平均值并且推送回客户端。服务请求的总响应时间，即从客户端发出服务请求到收到系统推送的服务响应的间隔时间(记为 T)由服务请求从客户端发送到系统的传输时间(记为 T_1)，系统计算响应结果的计算时间(记为 T_2)和系统推送服务响应

图 4 云端物联网大数据管理系统的原型

Fig. 4. The prototype of cloud-side IoT big data management system

到客户端所需的推送时间三部分构成，即 $T=T_1+T_2+T_3$。分析可知 T_1 和 T_3 由通讯信道的性能决定，而 T_2 由云端物联网大数据管理系统的实时计算性能决定，因此 T_2 是反映云端系统性能的关键部分，因此测试记录随车辆总数的不断增长，每次服务响应中 T_2 大小的变化情况。

(3) 测试目的和评价指标：测试目的为用 T_2 来评估系统对云端实时响应服务请求的响应可行性和性能，评价指标为随着车辆总数的不断增长，T_2 量级的增长情况(以秒为单位)。

(4) 测试数据集：测试数据集采用南京市某出租车公司公开的车辆状态数据集，数据集大小为 100 GB，以不同的车辆数为参数来划分数据集作为每次测试的测试数据。

(5) 结果与分析：重复执行测试用例，统计随着车辆总数不断增加时 T_2 的增长情况，得出车辆总数和 T_2 的关系图，如图 5 所示。分析可知，车辆每 3 秒上载一次位置和速度数据，每辆车 3 分钟内上载 60 次数据，车辆数越多，云端系统需要筛选和计算的数据记录数就越多，从而 T_2 越大。车辆数为 1000 辆时，3 分钟内上载的数据记录为 6 万条，T_2 为 0.71 秒；车辆数增加到 50 万辆时，3 分钟内上载的数据记录为 3000 万条，T_2 仅为 4.87 秒，由此评估出本系统对于云端实时响应服务请求的响应具有较高的实时性。

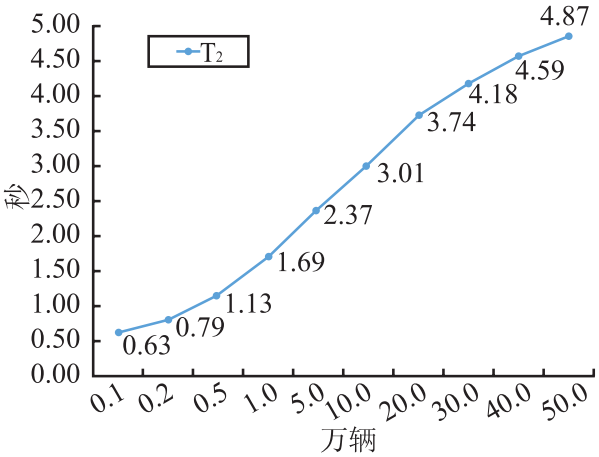

图 5　车辆总数与 T_2 的关系图

Fig. 5.　The variation of T_2 with increment of vehicles

5.2　云端大数据分析与挖掘服务请求

本测试案例对南京市某出租车公司公布的出租车载客历史数据进行载客区域密度分布来评估云端物联网大数据管理系统对云端大数据分析与挖掘服务请求的响应性能。

(1) 测试环境搭建：测试环境同 5.1。

(2) 测试用例设计：对南京市某出租车公司公布的出租车载客历史数据进行载客区域密度分布分析，其中一条载客数据由载客时间、载客地点(经度和纬度表示)、出租车车牌号等 3 个字段标识。统计随着历史载客数据集大小不断增加时所需分析时间的增长情况。

(3) 测试目的与评价指标：测试目的为用云端系统分析南京市出租车载客区域密度分布所需的分析时间来评估云端系统对于大数据分析与挖掘服务请求的性能表现，评价指标为随着数据集大小的增长，所需的分析时间量级的增长情况。

(4) 测试数据集：测试数据集为南京市某出租车公开的出租车载客历史数据，数据集大小为 50 GB。

(5) 结果与分析：采用数据集的子集作为每次测试的测试数据集，大小从 5 GB 增加到 50 GB，统计随着历史载客数据集大小不断增加时所需分析时间的增长情况，如图 6 所示。由图 6 可知，对于 10 GB 量级的大数据分析与挖掘任务，本系统具有较高的性能。测试案例分析结果的可视化表示如图 7 所示，其采用的分析数据集大小为 10 GB。图 7 所示为在某段时间内南京市某出租车公司的上客区域密度在 Google 卫星地图上的分布，其中彩色的点表示在对应地点有出租车上客记录，图中点呈现黑色点块的部分由大量密集的彩色点叠加形成，表示在对应区域的客流量比较密集，从而给出租车公司和司机以指导，合理规划车辆分布，方便客户的同时提高出

租车公司的收益。

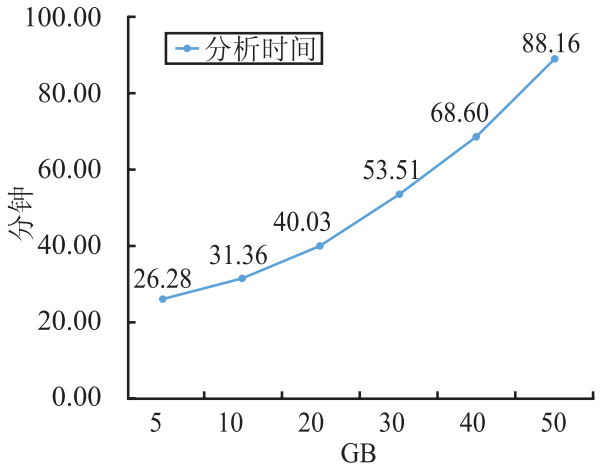

图 6 数据集大小和分析时间关系图

Fig. 6. The variation of the analysis time with increment of the data set

图 7 载客区域密度分布图

Fig. 7. Passenger carrying density distribution in region

6 相关研究

海云协同是一个创新的实验模型，目前关于这方面的研究并不广泛：孙凝辉等[17]从发展的角度讨论了海计算模型；王姣龙[18]讨论了物联网、云计算和海计算之间的关系；葛敬国等[19]提出海云试验环境管控与服务体系的一种总体架构；所有这些研究都只是给出了海云协同模型的一个抽象概念和架构，并没有详细的设计和实现细节，本文给出了海云协同模型的一个详细的架构和实现方案，包括海端计算系统的详细设计方案、云端物联网大数据管理系统的设计和实现方案以及海云协同模型的协同机制。关于物联网大数据管理的研究，有大量的研究工作在学术界和工业界开展：黄哲学等[20]讨论了海云数据系统的关键技术和系统的研制；Zaslavsky 等[21]讨论了物联网系统架构，大规模传感网应用以及基于云计算的物联网传感数据的存储和处理等问题的研究；Laurila 等[22]讨论了移动计算中大数据处理的挑战性问题；Ma 等[23]提出了一种新型的物联网大数据处理的索引框架；亓开元等[24]讨论了针对高速数据流的大规模实时处理方法；Ding 等[25]提出了一种管理和检索海量异构传感数据的数据库集群框架。

许多大型公司例如 Facebook、Yahoo!和 Twitter 等致力于研发大数据管理和分析开源项目。与 Apache Hadoop 相关的项目例如 Apache Pig、Apache Hbase 和 Apache ZooKeeper 等都致力于分布式大数据处理。Storm[26]是由 Twitter 发起的一个针对分布式实时流计算的开源项目。对于大数据挖掘问题，有一些优秀的开源项目。Apache Mahout[27]是一个基于 Apache Hadoop 的可扩展数据挖掘和机器学习开源项目；MOA[28]是一个开源的实时流数据挖掘软件；Vowpal Wabbit[29]是一个开源的可扩展机器学习项目，由 Yahoo! 发起。

7 结 论

本文提出了一种面向物联网大数据管理的创新型解决方案：面向物联网大数据管理的海云协同模型。在第 2 部分讨论了海云协同模型的整体架构和协同机制以及三种不同类型服务请求的响应机制；第 3 部分详细讨论了海云协同模型中海

端系统的设计和实现方案；第 4 部分讨论了海云协同模型中云端物联网大数据管理系统的系统架构和实现方案并且介绍了基于此系统架构开发的原型系统；最后在第 5 部分，用道路拥堵状况实时获取测试案例和南京市出租车载客区域密度分布分析测试案例评估了云端物联网大数据管理系统的实践可行性和性能表现。案例分析结果表明本文提出的解决方案具有实践可行性和较好的性能表现。

参 考 文 献

[1] Atzori L, Iera A, Morabito G. The internet of things: a survey [J]. Computer Networks, 2010, 54(15): 2787-2805.

[2] Francis D. On the origin(s) and development of the term ‘big data’ [EB/OL]. University of Pennsylvania. http://economics.sas.upenn.edu/pier/working-paper/2012/origins-and-development-term-%E2%80%9Cbig-data.

[3] Francis D. ‘Big data’ dynamic factor models for macroeconomic measurement and forecasting [C] // Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society, 2003: 115-122.

[4] Luo HY, Zerfos P, Kong JY, et al. Self-securing ad hoc wireless networks [C] // Proceedings of IEEE Symposium on Computers and Communications, 2002: 548-555.

[5] Sadao O, Naoto K, Davis P. Break throughs in large-scale ad hoc wireless networking and application for vehicle safety [C] // Proceedings of 7th International Conference on Mobile Data Management, 2006: 88.

[6] Yan XP, Zhang H, Wu CZ. Research and development of intelligent transportation systems [C] // Proceedings of 11th International Symposium on Distributed Computing and Applications to Business, 2012: 321-327.

[7] Kay R, Friedeman M. The design space of wireless sensor networks [J]. Wireless Communications, 2004, 11(6): 54-61.

[8] Agrawal R, Srikant R. Fast algorithms for mining association rules [C] // Proceedings of 20th International Conference on Very Large Data Bases, 1994: 321-327.

[9] Han JW, Pei J, Yin YW. Mining frequent patterns without candidate generation [C] // Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, 29(2): 1-12.

[10] Amazon Web Service, Inc. Amazon Simple Storage Service Developer Guide [EB/OL]. http://awsdocs.s3.amazonaws.com/S3/latest/s3-dg.pdf.

[11] Chang F, Dean J, Dhemawat S, et al. Bigtable: a distributed storage system for structured data [J]. ACM Transactions on Computer Systems, 2008, 26(2): 4.

[12] Borthakur D. HDFS architecture guide [EB/OL]. http://hadoop.apache.org/common/docs/current/hdfs/_design.pdf.

[13] OpenStack. Open Source Software for Building Private and Public Clouds [EB/OL]. http://www.openstack.org.

[14] Dean J, Ghemwat S. MapReduce: simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107-113.

[15] Apache Hadoop [EB/OL]. http://hadoop.apache.org.

[16] Rudack M, Meincke M, Lott M. On the dynamics of ad hoc networks for inter vehicle communication (IVC) [C] // Proceedings of the ICWN, 2002.

[17] 孙凝辉, 徐志伟, 李国杰. 海计算: 物联网的新型计算模型 [J]. 中国计算机学会通讯, 2010, 6(7): 39-43.

[18] 王娇龙. 物联网与云计算、海计算的关系 [J]. 物联网技术, 2012, 2(2): 15-19.

[19] 葛敬国, 唐海娜, 鄂跃鹏, 等. 海云创新试验环境管控与服务系统总体设计 [J]. 网络新媒体技术, 2012, 1(6): 45-51.

[20] 黄哲学, 曹付元, 李俊杰, 等. 面向大数据的海云数据系统关键技术研究 [J]. 网络新媒体技术, 2012, 1(6): 20-26.

[21] Zaslavsky A, Perera C, Georgakopoulos D. Sensing as a service and big data [C] // Proceedings of the International Conference on Advances in Cloud Computing (ACC), 2012.

[22] Laurila JK, Gatica-Perez D, Aad I, et al. The mobile data challenge: big data for mobile computing research [C] // Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing, 2012: 1-8.

[23] Ma YZ, Rao J, Hu WS, et al. An efficient index for massive IOT data in cloud environment. [C] // Proceedings of the 21st ACM international conference on Information and knowledge management, 2012: 2199-2133.

[24] 亓开元, 赵卓峰, 房俊, 等. 针对高速数据流的大规模数据实时处理方 [J]. 计算机学报, 2012, 35(3): 477-490.

[25] Ding ZM, Xu JJ, Yang Q. SeaCloudDM: a database cluster framework for managing and querying massive heterogeneous sensor sampling data [J]. The Journal of Super Computing, 2012: 1-25.

[26] Storm [EB/OL]. http://storm-project.net.

[27] Apache Mahout [EB/OL]. http://mahout.apache.org.

[28] MOA [EB/OL]. http://moa.cms.waikato.ac.nz.

[29] Vowpal Wabbit [EB/OL]. http://hunch.net/~vw.