

# 面向家庭物联网的云计算架构\*

陈亮<sup>a</sup>, 仇晶<sup>b</sup>, 朱有产<sup>b</sup>, 顾雪平<sup>a</sup>

(华北电力大学 a. 电气与电子工程学院; b. 控制与计算机工程学院, 河北 保定 071003)

**摘要:** 针对目前家庭物联网智能终端处理方式单一, 生成模式非动态和服务模式人性化程度低等问题, 提出一种面向家庭物联网环境的云计算架构体系。基于 MINA 框架和 JMS 消息服务设计出家庭物联网的消息管理机制, 根据 HBase 数据库设计家庭物联网的数据存储模型以实现数据存储分区, 采用 MapReduce 技术实现家庭物联网的数据分析及数据挖掘, 从而构建出完备的家庭物联网云框架。最后数据存储、数据分析和挖掘的性能测试结果证明了云计算框架的有效性和高效性。

**关键词:** 云计算; 家庭物联网; 数据存储; 数据分析与挖掘

**中图分类号:** TP393    **文献标志码:** A    **文章编号:** 1001-3695(2013)12-3686-04

doi:10.3969/j.issn.1001-3695.2013.12.043

## Cloud computing architecture oriented household Internet of things

CHEN Liang<sup>a</sup>, QIU Jing<sup>b</sup>, ZHU You-chan<sup>b</sup>, GU Xue-ping<sup>a</sup>

(a. School of Electrical & Electronic Engineering, b. School of Control & Computer Engineering, North China Electric Power University, Baoding Hebei 071003, China)

**Abstract:** Considering the problems of the current household Internet of things intelligent terminals—single processing mode, non-dynamic generation mode and low level humanized service mode, this paper proposed a cloud computing architecture for the household Internet of things. The message management mechanism of the proposed architecture was based on MINA framework and Java message service. And its data storage module was based on HBase, achieving data segmentation. Besides, it applied MapReduce technology to realize data analysis and data mining. Accordingly, established a complete cloud computing framework for the household internet of things. At last, it tested the performances of data storage, data analysis and data mining of proposed architecture. The results prove its effectiveness and efficiency.

**Key words:** cloud computing; household Internet of things; data storage; data analysis and mining

## 0 引言

进入 21 世纪, 物联网技术的发展, 将进一步有助于智能电网的实现, 如电网与用户间的智能互动<sup>[1-3]</sup>。目前家庭物联网环境的研究更多是针对前端感知层, 然而物联网感知信息的价值提升和效能发挥, 取决于后端感知信息的集成处理与系统管控等方面, 即物联网长链的后端。当前家庭物联网对于家庭环境的感知能力及控制能力都有限, 只是按照固定预设模式, 不能根据不同用户动态地生成人性化的服务。因此需要建设一个后端处理层, 对采集的各种家庭物联网环境数据进行全面科学的分析及挖掘。而这就需要一种高效的通信管理策略、一个能容纳海量数据的存储平台以及一个强大的计算平台。

在数据存储技术上, 无论是传统的磁盘阵列存储还是 NAS (network attached storage, 网络附属存储) 或 SAN (storage area network and SAN protocols, 存储区域网络及其协议) 存储, 都存在着扩展性差、数据维护复杂等问题<sup>[4,5]</sup>。新兴的云存储技术超越了传统的数据存储模式, 可提供海量存储空间, 并支持灵

活扩展、高性能访问的文件共享存储。

人们对机器智能化的需求也促进了数据分析与数据挖掘技术的发展和研究, 特别是针对海量数据集的并行数据处理方法的研究。传统的数据分析是基于关系数据模型的, 其并行操作是通过昂贵的硬件 (如大规模并行处理 (MPP) 系统或对称多处理 (SMP) 系统) 来实现的, 其成本高, 扩展性差, 不能适应对海量非结构化数据分析。基于 Hadoop 的 MapReduce 并行分析, 支持面向列的数据模型, 灵活可扩展, 目前主要的算法实现有 Mahout 算法框架<sup>[6]</sup>。

本文采用云计算技术 Hadoop 框架<sup>[7]</sup>能够实现海量数据存储及超大计算能力。Hadoop 框架主要是由 HDFS (hadoop distributed file system)、MapReduce 和 Hbase 组成。其中: HDFS 是 Hadoop 的分布式文件系统; MapReduce 是并行计算框架, 适用于对海量数据的处理过程; Hbase 是类似 Google BigTable 的分布式 NoSQL 列数据库, 建立在 HDFS 之上, 提供高可靠性、高性能、列存储、可伸缩、实时读写的大数据库系统, 能够为海量数据的存储以及分析计算提供解决方案。

由于大量家庭物联网智能终端与云计算后台处于长连接

收稿日期: 2013-04-15; 修回日期: 2013-05-21    基金项目: 国家教育部大学生创业实践项目 (11062)

**作者简介:** 陈亮 (1989-), 男, 山东济南人, 硕士研究生, 主要研究方向为智能技术在电力系统中的应用 (raymond05@yahoo.cn); 仇晶 (1990-), 男, 陕西富平人, 硕士研究生, 主要研究方向为云计算的应用及安全; 朱有产 (1963-), 男, 浙江浦江人, 教授, 硕导, 博士, 主要研究方向为管理信息系统、计算机网络及应用; 顾雪平 (1964-), 男, 河北石家庄人, 教授, 博导, 博士, 主要研究方向为智能技术在电力系统中的应用。

状态,同时还要保证两者之间通信的安全性、可靠性以及实时性,所以在解决家庭物联网智能终端与云计算后台的消息管理问题上,采用将家庭物联网环境中的消息分类的方法,将其分为异步消息和同步消息两部分,分别使用 JMS 和 MINA 框架来完成。云后台知识库是云后台智能化标准的最直接体现,该知识库的知识完整度、准确度最直接地影响家庭物联网服务的品质。因此采用了适合于云计算的一套机器学习算法来构造云计算后台的知识库,主要为 Mahout 算法框架,通过聚类算法、分类算法<sup>[6]</sup>等算法的定时数据分析来不断更新知识库。

自从云计算诞生以来,云安全问题就受到了人们的广泛关注<sup>[8]</sup>,传统的数据访问控制策略存在着不能完成对数据访问的细粒度访问控制和避免特权用户非法访问用户的敏感数据等问题,因此基于 ABE 属性<sup>[9]</sup>加密算法的访问控制策略的研究逐渐成为热点。本文采用基于 CP-ABE 加密算法<sup>[10,11]</sup>的访问控制策略,通过引入由数据属主独立控制的许可属性,构建不同属性域的 CP-ABE 方案,以解决云存储环境下的数据安全问题。

## 1 面向家庭物联网的云计算架构设计

### 1.1 云计算架构

针对家庭物联网环境的要求,给出了如图1所示的云计算架构,主要包括系统维护模块、安全管理模块、功能层、业务逻辑层、数据处理层、数据存储层、数据接入层等。

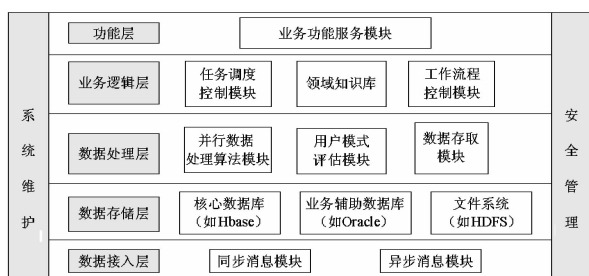


图1 面向家庭物联网的云计算架构

系统维护模块主要负责该架构系统的日志审计、异常处理、数据备份等工作。

安全管理模块主要包括数据加密和访问控制,主要作用是保证云架构系统的数据安全。

功能层主要包括业务功能服务模块,包含知识下载、模式更新、数据查询、消息决策等,是云后台对外提供的主要接口。

业务逻辑层主要包括任务调度控制模块、领域知识库、工作流程控制模块等。任务调度控制模块根据不同的处理请求,调用不同的业务处理或方法模块;工作流程控制模块对业务状态进行监控、管理,可将具体的信息参数返回给本层的任务调度控制模块。

数据处理层,主要包括并行数据处理算法模块、用户模式评估模块、数据存取模块等。并行数据处理算法模块为数据分析、数据挖掘的各种任务提供并行算法,主要包含一些基于云计算并行计算 MapReduce 的算法库;用户模式评估模块对产生的模式进行评估,符合用户要求的结果存入领域知识库,领域知识库可以辅助业务控制逻辑指导数据处理过程;数据存取模块根据不同的需要从数据存储层调取数据或者写入数据。

数据存储层,主要包括核心数据库、业务辅助数据库、文件

系统等。其中核心数据库采用 HBase 数据,负责系统主要元数据的存储;业务辅助数据库则可以采用 Oracle 等关系数据库来辅助系统的业务快速处理;文件系统采用 HDFS 分布式文件系统。

数据接入层主要包括同步消息模块和异步消息模块,主要负责智能终端与云系统的通信管理及消息处理,还包括一些对消息的预处理以方便数据存储层的存储和进一步的应用。其中异步消息部分是基于 JMS 异步消息服务,主要接收智能终端发来的非异常性及非指令性消息,其目的是解决大量智能终端与云端频繁通信而带来的负载问题。而同步消息处理部分则是基于 MINA 框架的消息通信管理,先在智能终端与云端之间建立合法会话,然后传送异常性及指令性消息,并实时调用云后台分析,返回操作决策,保证消息的实时处理;同时对智能终端的更新也采用这种方式。

### 1.2 云计算架构的工作过程

图2为本文云计算架构的工作过程示意图,这里将各种数据采集设备以及用户的请求动作统称为采集器。智能终端根据采集器的消息,将消息进行简单的预处理并分类为不同的传输方式和消息级别。

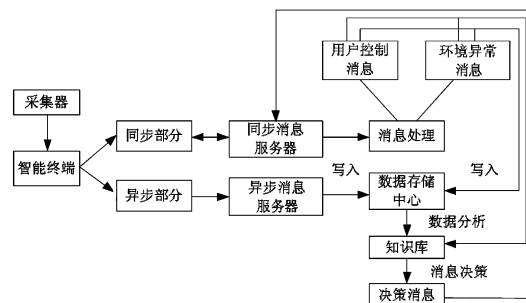


图2 面向家庭物联网的云计算框架的工作过程

以下为具体的通信过程:

#### 1) 同步消息传送

- 智能终端将消息发送到同步消息服务器;
- 同步消息服务器对消息进行甄别,分为用户控制消息和环境异常消息;
- 同步消息服务器将消息写入云后台数据存储;
- 同步消息服务器根据消息的级别分别调用数据分析方法对消息进行知识匹配,得出智能决策消息;
- 同步消息服务器将决策消息发回对应智能终端。

#### 2) 异步消息传送

- 智能终端将消息发送到异步消息服务器;
- 异步消息服务器对消息预处理之后写入云后台数据存储。

云后台服务器定期对存储数据进行分析挖掘,从而对知识库进行更新,保证消息决策的准确性。

## 2 技术说明

### 2.1 数据存储

该系统数据存储采用 Hadoop 框架中 HDFS 分布式文件系统,其能提供高容错、高吞吐量的数据访问;系统采用 HBase 数据库,它是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统,本文基于 HBase 数据库建立家庭物联网环境数据存储

模型,开发了适合智能家居物联网的 HomeTable。HomeTable 的数据模型见表 1。每一个 HomeTable 都是一个稀疏的、分布式的多维有序图,按行键值、列键值和时间戳建立索引。表的每个行都是动态分区,每一个分区称为 Tablet,是调整数据分布和负载均衡的最小单位。HomeTable 的列组是将多个相关列并为一小组,以提高存取和读取的效率。HomeTable 的时间戳的类型是 64 位整数,通过时间戳来进行实时信息的更新,以实现实时状态的监测。表 1 是 HomeTable 的数据模型,通过 Tablet 和列组两大特性提高访问效率和伸缩性。

表 1 HomeTable 的数据模型

| Tablet  | Rowkey | Dev_Kind | 标志码 (Dev_Id) |            | 状态信息 (State) |      | 操作指令 (Command) |      |           |
|---------|--------|----------|--------------|------------|--------------|------|----------------|------|-----------|
|         |        |          | RFID 标签      | UPnP IP 标志 | 设备状态         | 电网状态 | 查询指令           | 控制指令 | 联动方式      |
| Tablet1 | 510501 | 智能空调     | *****        | *****      | 开启           | 低谷   | 无              | 启动   | 关联 510504 |
|         | 510502 | 电热水器     | *****        | *****      | 开启           | 低谷   | 有              | 无    | 无         |
| Tablet2 | 510503 | 洗衣机      | *****        | *****      | 开启           | 低谷   | 无              | 启动   | 无         |
|         | 510504 | 门窗控制     | *****        | *****      | 控制关闭         | 低谷   | 无              | 无    | 关联 510501 |

HomeTable 通过行关键字的字典顺序来组织数据,即支持 Rowkey 作为唯一主键,并且可被用于索引和查询。表的每个行都是动态分区 (Tablet),是调整数据分布和负载均衡的最小单位。通过选择合适的行关键字,在访问数据时有效利用数据的位置相关性。更重要的是,在表 3 存储的数据增加的情况下,HomeTable 会将这个表切为多个 Tablet,并让这些 Tablet 分布到不同的服务器上,而每个 Tablet 都负责一定范围的列。例如在表 1 中,Tablet2 负责 Rowkey 为 510502 至 510503 的数据,可见 Tablet2 储存的是普通家电的信息。

HomeTable 的列组是将多个相关列并为一小组。存放在一列组下的所有数据通常是同一类型。因为每个表格都有成百上千列,而大多数查询只需得到其中少数几个列。如在表 1 中,“RFID 标签”和“UPnP-IP 标志”都属于“标志码 (Dev\_Id)”这个列组,“设备状态”和“电网状态”都属于“状态信息 (State)”这个列组,“查询指令”“控制指令”和“联动方式”都属于“操作指令 (Command)”这个列组。这样的好处是,将这些相关的列放在一起存储,不仅能提高存取效率,而且能避免读取过多的列。

在 HomeTable 中,表的每一个数据项都可以包含同一份数据的不同版本,不同版本的数据通过时间戳来索引。HomeTable 的时间戳类型是 64 位整数,HomeTable 为时间戳赋值,用来表示精确到毫秒的实时时间。数据项中,不同版本的数据按照时间戳倒序排列,从而使得最新版本最先被读取,而且系统能根据时间戳来做 GC (garbage collection, 垃圾回收)。如在表 1 中,列组“状态信息 (State)”的数据项就是通过时间戳来进行实时信息的更新,以实现实时状态的监测。

为了保证家庭物联网数据存储安全,同时兼顾系统性能需求,采用数据加密技术和签名技术对敏感数据进行加密存储,保障数据完整性和机密性,主要包括 AES、SHA-256 等算法。

## 2.2 数据分析与挖掘

家庭物联网云计算架构其核心目标就是为家庭物联网智能终端提供高性能、高可靠的后台数据分析处理方法,因此在

家庭物联网数据处理部分需要包含机器学习算法以及常规的数据分析算法来为智能终端作消息决策。

### 2.2.1 MapReduce 的处理流程

MapReduce 处理流程的流程如图 3 所示,步骤如下:

- 对输入数据源进行切片;
- master 调度 worker 执行 map 任务;
- worker 读取输入源片段;
- worker 执行 map 任务,将任务输出保存在本地;
- master 调度 worker 执行 reduce 任务,reduce worker 读取 map 任务的输出文件;
- 执行 reduce 任务,将任务输出保存到 HDFS。

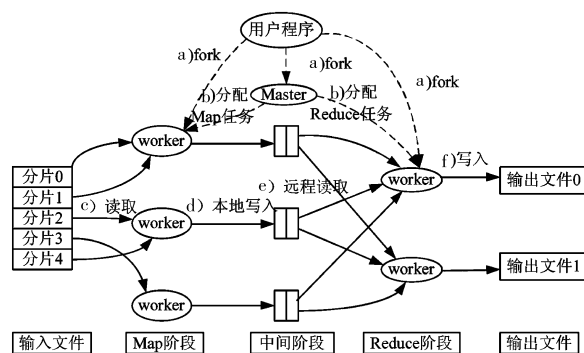


图 3 MapReduce 的处理流程

### 2.2.2 决策树处理法

决策树是一种依托于策略抉择而建立起来的树,能够辅助数据分析等处理过程。

下面以家庭物联网云计算后台领域知识库中的煤气知识决策树为例说明:在领域知识库中预先存储了处理煤气浓度超标的决策树,在检测到环境中煤气浓度超标时就会根据该决策树进行后续处理,从而保证家庭生活环境的安全。其决策树如图 4 所示。

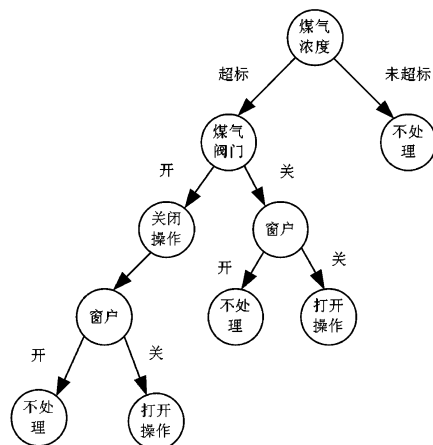


图 4 煤气浓度超标决策树

### 2.2.3 Mahout 算法框架

Mahout 算法框架中的算法涵盖:分类算法、聚类算法、关联规则挖掘算法、推荐/协同过滤算法、降维/维约简算法等,并且由于其本身就是基于 Hadoop 实现的,所以很多算法都已经有 MapReduce 模式的实现,这样就大大提高了算法可处理的数据量和数据处理的性能。

图 5 为 Mahout 中 K-means 聚类算法数据分析过程。虽然 K-means 简单且高效,但它存在一定问题,首先 K 值 (即簇的数

量)是人为确定的,在对数据不了解的情况下,很难给出合理的  $K$  值;其次初始簇心的选择是随机的,若选择到了较孤立的点,会对聚类的效果产生非常大的影响。因此与 Canopy 算法配合,进行初始化,确定簇数以及初始簇心。

### 2.3 家庭物联网云存储系统访问控制策略

访问控制是实现用户数据机密性和进行隐私保护的重要手段,目前有很多 CSP (cloud storage provider, 云存储服务提供商)都提供简单的访问控制功能,这些安全机制完全依赖于 CSP 服务器端,但不能确保 CSP 在复杂的网络环境和多变的商业利益之中能够始终保证安全机制的有效性。因此本文在家庭物联网云存储系统中使用基于 CP-ABE 加密算法的访问控制策略。通过引入由数据属主独立控制的许可属性,构建不同属性域的 CP-ABE 方案,能够防止云存储系统特权用户的内部攻击,使数据属主能完全控制其他共享用户对其云数据的访问。

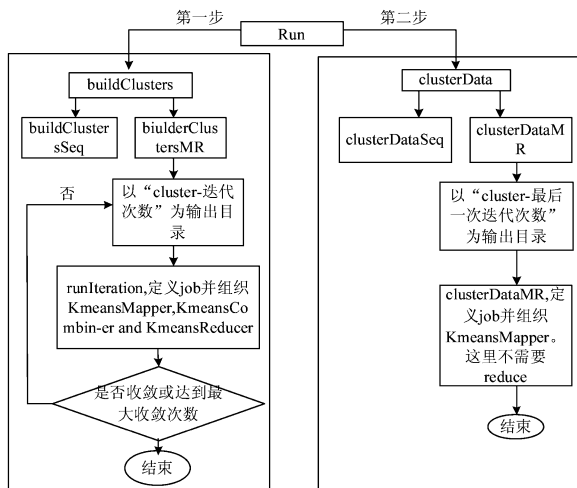


图5 K-means 数据分析过程

### 3 性能测试分析

实验环境:搭建 Hadoop 机群测试平台,共有 15 台机器,均为同等配置(内存 2 GB,硬盘空间 160 GB),操作系统为 Ubuntu 10.04, Hadoop 0.20.2, HBase 0.90.5。

图6为对家庭物联网云存储系统的读写能力进行的测试,分别以 1、5、10、50、100、222、485、815 MB 数据作为测试数据。对测试结果进行分析能够得出:该系统在处理小于 50 MB 文件时,读写速度变化幅度较大,而对于大于 50 MB 的文件,则能够维持在一个基本稳定的速度,即写入速度 7.30 MBps,读取速度 4.42 MBps。

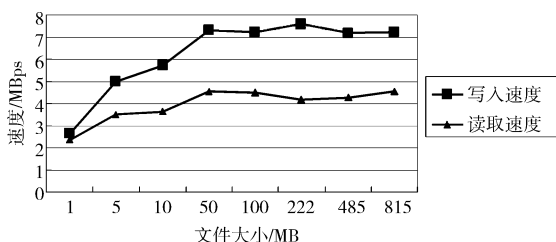


图6 云存储系统读写性能

图7为对 Canopy + K-means 的 MapReduce 并行实现与单机实现的性能测试。分别以 20、200、2000 MB 样本数据作为测

试数据集,以 1、3、4、6、8、10、12 个节点为测试集群。可以看出,在处理小数据集时,单机的性能比并行性能优越;但在处理大数据集时,MapReduce 并行集群运算的性能优越性表现得十分明显,而这也是 MapReduce 设计的初衷,也能够很好地满足家庭物联网环境对数据分析和数据挖掘的效率需求。

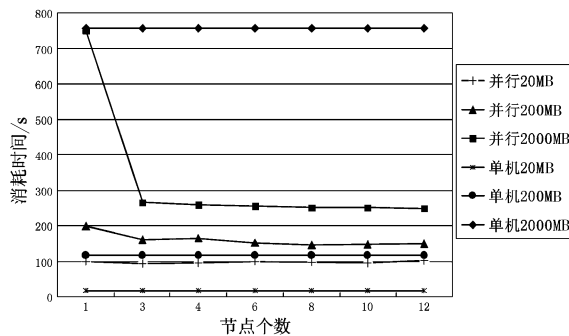


图7 Canopy + K-means 的 MapReduce 并行性能与单机性能测试

### 4 结束语

针对家庭物联网环境数据处理能力不足,提供的服务不够智能化、人性化的问题,本文基于 MINA 框架和 JMS 消息服务设计出家庭物联网的消息管理机制,根据 HBase 数据库设计家庭物联网的数据存储模型,实现了数据存储分区,采用 MapReduce 技术实现了家庭物联网的数据分析及数据挖掘,从而构建出完备的家庭物联网云框架。通过对本文云计算架构的数据存储、数据分析和挖掘的性能测试,证明了该框架的有效性,为家庭物联网的高效智能数据处理提供了一种创新可行的技术方案。

#### 参考文献:

- [1] FARHANGI H. The path of the smart grid[J]. IEEE Power and Energy Magazine, 2009, 8(1): 18-28.
- [2] 朱洪波, 杨龙祥, 于全. 物联网的技术思想与应用策略研究[J]. 通信学报, 2010, 31(11): 2-9.
- [3] 杨震. 物联网发展研究[J]. 南京邮电大学学报, 2010, 12(2): 1-10.
- [4] 王宇明. 磁盘阵列存储技术简析[J]. 现代电视技术, 2006, 21(7): 140-143.
- [5] 段剑弓. 存储系统 NAS 和 SAN 的差异和统一[J]. 计算机应用研究, 2004, 23(12): 94-97.
- [6] OWEN S, ANIL R, DUNNING T, et al. Mahout in action[M]. New York: Manning Publications Co, 2010: 25-30.
- [7] 崔杰, 李陶深, 兰红星. 基于 Hadoop 的海量数据存储平台设计与开发[J]. 计算机研究与发展, 2012, 49(S1): 12-18.
- [8] 冯登国, 张敏, 张妍, 等. 云计算安全研究[J]. 软件学报, 2011, 22(1): 71-83.
- [9] BETHENCOURT J, SAHAI A, WATERS B. Ciphertext-policy attribute-based encryption[C]//Proc of IEEE Symposium on Security and Privacy. Washington DC: IEEE Computer Society, 2007: 321-334.
- [10] 刘帆, 杨明. 一种用于云存储的密文策略属性基加密方案[J]. 计算机应用研究, 2012, 29(4): 1452-1456.
- [11] 孙国梓, 董宇, 李云. 基于 CP-ABE 算法的云存储数据访问控制[J]. 通信学报, 2011, 32(7): 146-152.