

# 建筑能耗监测大数据存储与服务框架研究

张云翼, 林佳瑞, 康凯, 张建平

(清华大学, 北京 100084)

**【摘要】** 建筑业信息化的不断发展使数据量呈现爆发式增长, 而大部分建筑能耗数据仍储存于关系型数据库中, 大数据体量的增加将超出其存储极限。本文针对建筑及其机电系统标准化大数据进行分析, 建立了实体-关系, 设计了静态数据及动态数据模式, 建立了基于 Hadoop 的云端数据库框架对海量数据进行标准化处理与存储, 并提供多层次应用服务对数据进行分析处理。工程实例验证表明, 本文提出的存储与服务技术可在正常使用状态下满足数据的管理与分析需求。

**【关键词】** 大数据; 存储管理; 动态服务

## 1 背景

近年来, 随着互联网等相关技术的快速发展, 各行业的信息呈现爆发式增长。信息量的快速增长使得相关人员可以通过分析和处理, 得到大量有价值的信息, 但数据量的爆炸使得传统处理方法已不能应对巨量信息<sup>[1,2]</sup>。这种在有限时间内难以用传统方法处理的数据, 通常被称为大数据。

在建筑行业, 信息化的不断发展也使得产生的数据量呈现爆炸式增长, 建筑信息模型 (BIM, Building Information Modeling) 技术的出现和发展, 使得建设领域的信息共享与交换变得更加方便, 也为大数据的应用提供了可能<sup>[3]</sup>。Natalija Koseleva 等<sup>[4]</sup>通过分析认为, 建筑及其机电监测数据完全符合大数据的四个特征, 然而大部分的能源数据仍然保存于传统的关系型数据库中, 而且随着数据量的不断增加, 传统关系型数据库将超过其存储能力的极限。大数据的体量将呈爆炸式增加, 对大数据的存储与分析方法都是一种挑战, 在这些技术上必须有所突破。

然而, 当前国内建筑能耗监测采用不同厂商提供的多种软件系统, 它们基本互不兼容, 除了少量数据外, 绝大多数数据无法实现有效融合集成管理, 不能用于大数据分析。本课题研究如何从将提取到的建筑及其机电系统静态标准化大数据和能耗监测系统的静态及动态大数据进行有机融合与集成, 以及如何支持服务应用。

## 2 存储管理技术

### 2.1 数据库总体设计

---

**【基金项目】** 国家重点研发计划项目“基于全过程的大数据绿色建筑管理技术与示范”(2017YFC0704200), 中国科协青年人才托举工程(2016QNR001), 清华大学(土木学院)-广联达 BIM 联合研究中心项目

**【作者简介】** 张云翼(1993-), 男, 博士研究生。主要研究方向为建筑信息模型与大数据。E-mail: yunyi2525@qq.com

基于调研结果和课题需求，数据库架构以 Hadoop 为基础，利用阿里云提供的计算服务进行搭建，支持接入与整合多源异构数据，进而对数据进行转化处理、统计、分析、挖掘，对数据隐藏的价值进行深度挖掘，最后对处理结果进行可视化展示或共享。平台分为数据源、接口层、数据集成层、分析存储层、服务层、应用层，如图 1 所示。

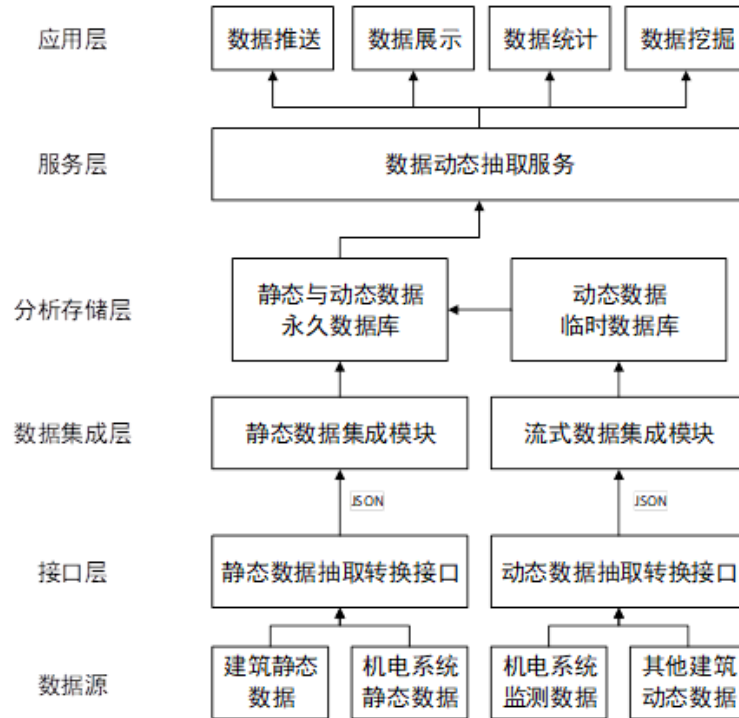


图 1 数据库总体设计

在接口层，主要从数据源获取数据，并通过静态及动态的数据抽取转换接口，将数据转换为标准化数据格式，通过预先约定的 JSON 等数据格式传输至数据集成层。

在数据集成层，分别利用静态和流式数据集成模块实现数据的接入和集成，同时满足短时间内对大规模数据进行抽取，和对生成频率高、时效性强的流式数据的处理。

在分析存储层，在海量数据场景下，单一服务器的架构无法满足存储或计算的需求，必须通过云架构实现分布式横向扩展。针对流式数据时效性强的特点，流计算模块可对流式数据进行实时计算，最终结果装载至永久数据库。

在服务层，根据应用需求从数据库中抽取数据并推送至应用层进行数据展示、统计、分析、挖掘等。其中数据展示可以对数据库中的数据进行查询和初步统计，查看数据基本情况，验证数据已根据预定模式集成。

## 2.2 数据模式分析

本文对数据库应存储的数据及其逻辑关系进行了梳理，结合现有工程数据的实际情况，构件以空间为核心的实体-关系。建筑、构件、机电系统等静态数据均与空间挂接，形成树

状关系；动态数据的原始数据、汇总数据、关键指标数据等同样与空间相挂接，以实现静态数据与动态数据的融合，支撑数据的深度挖掘。

### 2.3 静态数据模式设计

建筑基本信息以键值型数据为主，共设 3 个列族，**buildingCode** 列族存储建筑代码，**name** 列族存储建筑名称，**Property** 列族存储建筑的基本属性，该列族的不同列分别存储不同的属性，属性的具体命名存储于数据字典表中，行键设定为建筑代码。

表 1 建筑基本信息表

building 表					
RowKey	buildingCode	name	property		
			area	height	.....
建筑代码	建筑代码	建筑名称	建筑面积	建筑高度	.....

建筑构件、设备、空间信息通常存在上下级的包含关系，可组织为树形结构。例如，每一建筑可划分为不同的楼层，每一楼层又可划分为不同的区域，每一区域可进一步划分为不同的房间或空间等。考虑到这样的树形结构关系，每一节点可存储为键值型数据，而通过引用其父节点即可存储整个树形结构。以构件信息为例，可表达为如下结构。

表 2 建筑构件信息表

buildingEntity 表					
RowKey	parent	name	property		
			type	material	.....
建筑代码+构件代码	父节点行键	构件名称	构件类型	构件材质	.....

### 2.4 动态数据模式设计

目前对动态监测数据的存储共有三种主要方案<sup>[5]</sup>。其中，单点存储方案可以完整存储原始数据，且容易处理不同粒度的数据，但读写压力和存储空间较大。因此在本研究中，收集实时数据的临时数据库采用此种方案；而永久数据库则适合采用时间段存储的方式，即一条数据可存储一段时间内的多条数据，方便逐日、逐月、逐年进行统计分析，抽取所用数据进行进一步的分析挖掘。以建筑耗电总量为例，两种存储方案如下所示。

表 3 动态数据表（临时数据库）

rawData 表					
RowKey	buildingCode	time		data	
		date	stamp	item	value
建筑代码+时间+项目	建筑代码	日期	时间	项目	值

表 4 动态数据表（永久数据库）

monitorData 表					
RowKey	building	item	date	value	stats

				+0	+1	.....	sum	.....
建筑代码+时间+	建筑代码	项目	日期	0:00	1:00	.....	总和	.....
项目				监测值	监测值			

### 3 动态服务技术

#### 3.1 动态服务技术架构

从现有的建筑能耗监测案例和系统来看<sup>[6,7]</sup>，建筑及其机电系统标准化大数据服务应当满足的用户需求可以按照数据处理方式分为三类：数据查询、数据统计分析、数据挖掘等。应当提供的主要功能有：状态监测，提供各项机电设备的最新状态；异常报警，发现异常状态发出报警信息；分项统计，计算各项数据的均值极值等统计指标；图表与报表，使用表格和图片展示各项数据；节能控制，分析监测数据，调整建筑使用方式从而降低能耗等。

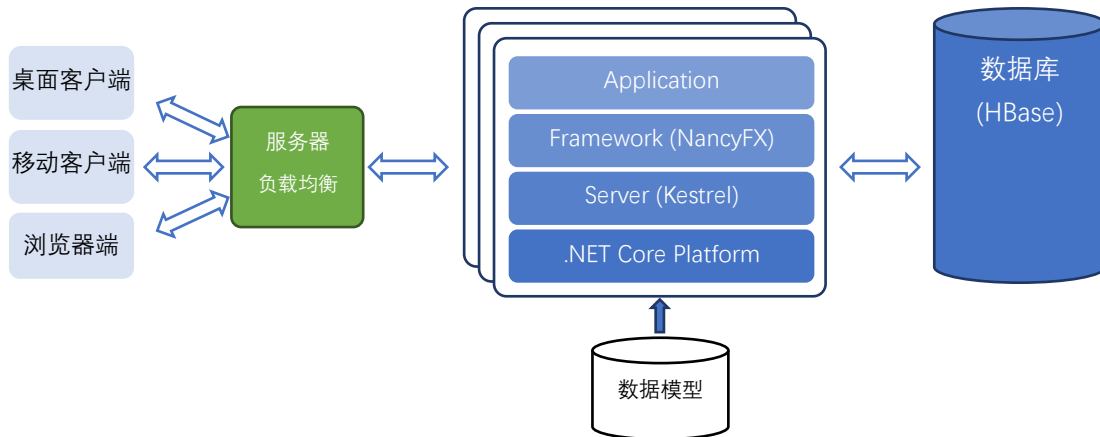


图2 系统技术选型

根据已经调研的现有的相关服务技术架构，本课题所采用的服务架构以微服务架构为主并参考其他架构进行设计。SOA架构的主要优势是相比传统架构提高了可重用性、敏捷性等，微服务架构更进一步地将系统分解为多个可独立运行的微服务，方便应用的自动化部署、迁移、维护和扩展，适合提供弹性云服务。

动态服务可选择.NET Core平台进行开发，它是一个开源、跨平台、高性能和轻量的应用程序开发框架，适合Web微服务的开发。在Web服务框架方面，NancyFX具有轻量化和动态注册路由的特点，开发接口简洁友好，功能支持较为全面，因此采用为Web服务框架。负载均衡功能可使用Nginx。应用容器引擎Docker提供了Linux虚拟化容器服务，是构建微服务系统常用的工具。内存数据库Redis等可提供高速的缓存服务。对于服务与数据库的连接方法，HBase提供了Java API、REST和Thrift协议的接口，本研究的动态服务程序通过Thrift 2协议的接口连接HBase数据库。

由于建筑信息动态服务涉及大量类型的数据，但不同类型之间具有相似性，因此可以

提取相同点，形成自动化工具，从数据模型生成大量服务代码，提高开发和维护效率，降低工作量和错误率。因此，可利用元编程思想处理这一问题，即以程序作为处理对象的编程技术。利用自动工具，从数据模型中获取数据类型信息，填入模板中生成服务代码。

### 3.2 动态服务接口设计

用户对监测数据的需求可分为三个层次。最底层的数据接口提供比较基础和原始的数据服务，能够进行比较简单的数据查询；中层的数据接口提供一定程度的聚合、条件查询服务，能够获取给定查询时间和空间范围内的统计量；高层的数据接口的目标是满足数据挖掘、统计分析等需求。

根据以上分析和数据服务需求，数据查询接口适合采用面向资源的 RESTful 接口设计方法，传输轻量、易解析的 JSON 格式数据。部分较为复杂的事务性操作应当参考 SOAP 的设计思路设计响应接口作补充。如果需要推送等交互方式，可使用 WebSocket 等技术实现双向通信服务。

## 4 工程实例验证

为测试数据库模式设计及读写效率，以某医院建筑一年的电能消耗实际数据为例，导入数据库进行试运行，该数据以小时为单位，共 8760 条。经测试，将一栋建筑全年的数据导入数据库仅需 4 秒，且包括与数据库建立连接的时间，因此在数据库投入运行后，预期可以满足写入效率的需要。

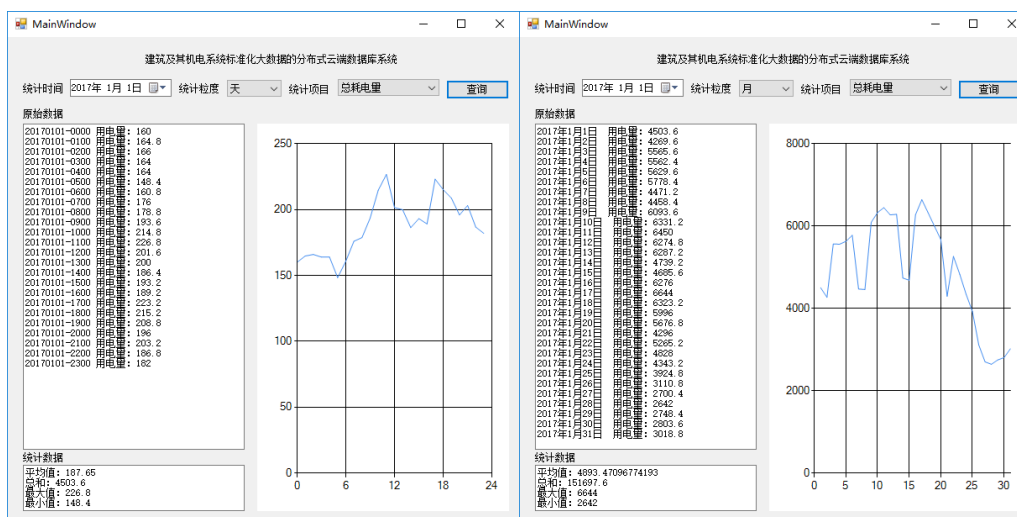


图 3 2017 年数据读写效率测试

图 3 中展示了 2017 年 1 月 1 日每小时的耗电量变化及 2017 年 1 月每日的耗电量变化。从图中可以看出耗电量的日变化呈现双峰趋势，即工作时间的耗电量大于非工作时间，午休时间耗电量略有降低；而月变化呈现周期性变化趋势，工作日耗电量明显大于休息日。

动态服务接入了上述数据库，并使用数据库已经导入的测试数据进行了一定的测试，例如图 4 中测试了 2017 年 1 月 1 日至 31 日期间的数据查询功能。动态服务程序接收到查询请求后，与数据库建立连接、提取数据、将数据序列化为 JSON 格式，返回给查询者。经过测试，服务程序按照预期返回了查询时间段内的数据，响应时间在 1 秒以内。

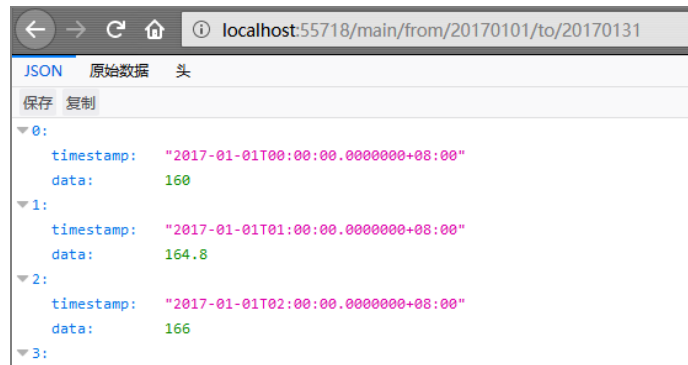


图 4 2017 年 1 月数据查询

## 5 结论

本文针对建筑及其机电系统标准化大数据进行分析，建立了实体-关系，设计了静态数据及动态数据模式，建立了基于 Hadoop 的云端数据库对海量数据进行标准化处理与存储，并提供多层次应用服务对数据进行分析处理。工程实例验证表明，本文提出的存储与服务技术框架可在正常使用状态下满足数据的管理与分析需求。

### 参考文献

- [1] Anand R. More Data Usually Beats Better Algorithms[EB/OL]. [2017]. <http://anand.typepad.com/datawocky/2008/03/moredata-usual.html>.
- [2] C. A. The end of theory: The data deluge makes the scientific method obsolete[J]. Wired magazine, 2008,7(16).
- [3] Eadie R B M O H. BIM implementation throughout the UK construction project lifecycle: An analysis[J]. Automation in Construction, 2013(36):145-151.
- [4] Koseleva N, Ropaite G. Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges[J]. Procedia Engineering, 2017,172:544-549.
- [5] Davila D, Ioannis B. Modelling, management, and visualisation of structural performance monitoring data on BIM: International Conference on Smart Infrastructure and Construction,
- [6] 陈海波, 胡云. 基于能耗计量监测工程实践及调研的大型公共建筑节能计量客户需求分析[J]. 建筑节能, 2010(12):71-73.
- [7] 储雪松. 城市建筑能耗监测平台构建及管理需求的实现[J]. 建筑节能, 2010(06):74-76.