

# 十三五建筑供暖空调解决方案后台数据库设计

## 数据库目的

汇总小米和建研院在38个示范工程收集的温湿度和能耗数据并以网站的方式予以展示。网站要求能够提供登陆功能。

## 数据概览

- 最核心的数据是温湿度和能耗，这些数据由项目下的设备收集所得。
- 不同的数据可能收集自不同设备。以小米平台的数据为例，温湿度和电耗分别由两个设备收集所得。
- 单个项目下会有多个测点。即使源自多个设备，一条完整的温湿度和电耗数据逻辑上归属于一个测点。
- 有的数据源会提供测点的名称，每个设备也会有设备所在房间的名称（小米有环科智控没有）。
- 单个设备记录的项目有所不同。
- 设备记录每五分钟一条从2017年持续到2018年。
- 不同设备具体所测量的值会有所不同。
- 环科智控：地图，项目信息（下面有很多测点），设备。
- 小米：住户（相当于测点），设备，缺少项目信息。

## 数据库表名称定义

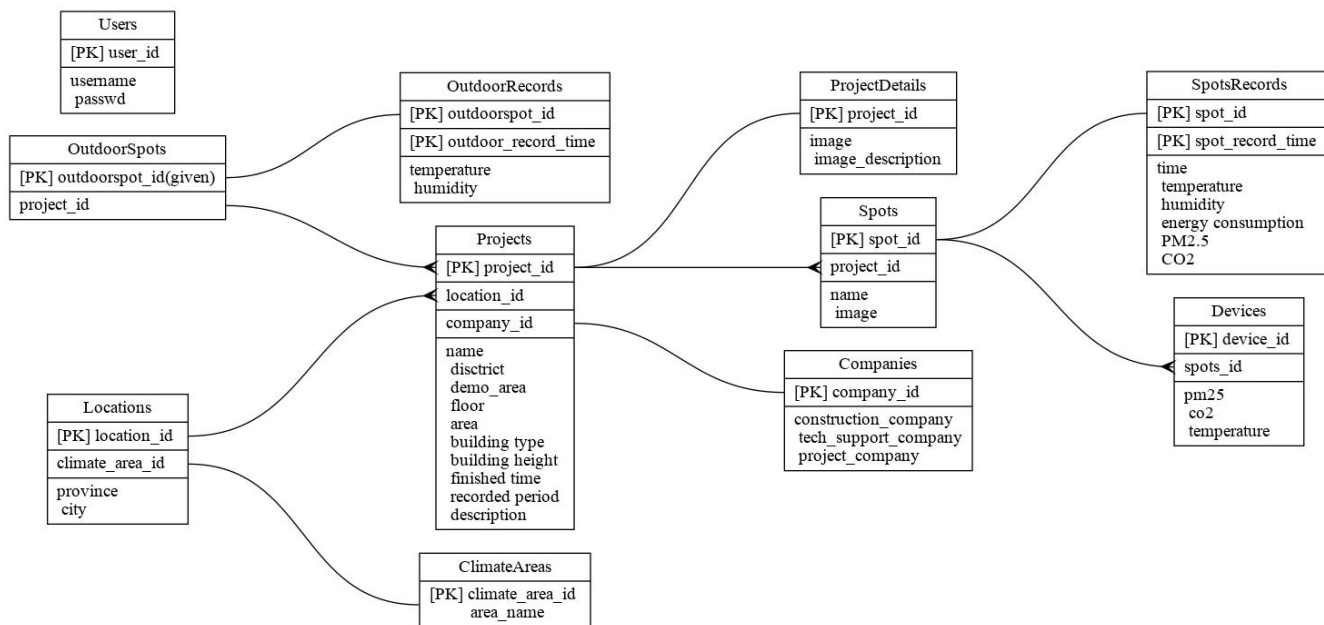
- 项目：38个示范工程之一
- 测点：项目下包含温度，湿度，电耗，pm2.5, co2的一个测量点。
- 设备：测点下具体的设备。

## 数据存储形和操作

- 数据存储在本服务器数据库。
- 服务器后台提供数据库python操作接口，封装对数据库的操作。
- 所有后续操作由调用api完成。可能的操作包括更新数据记录，常用的业务视图查询（需要讨论是否提供删除和修改逻辑）。

## 数据库table设计

- 用户信息单独列一表，与主业务表不关联。
- 单个项目可以有多个测点。
- 单个测点可以有多个设备。
- 项目的额外信息（图片）存在ProjectDetails表内。
- 核心数据（温湿度能耗）存在测点表内。
- 设备数据存在设备表内。



## 数据收集方案

- 待定，目前没有数据源的api。
- 首先在本地依照上述的表结构建立空库，再从数据源收录数据入数据库
  - 若有api，根据api制定具体的数据收集策略。
  - 若无api，需要从数据源的展示页面爬取（工作量大）

## 常用的业务视图

规定特定的业务视图方便网站后台调用以及后续的宣传。

- 项目粗视图
  - 项目id，项目名，项目公司，项目地点
- 项目细视图
  - 项目id，项目名，项目地点，项目公司，层高，面积，建筑型号，建筑高度，完工时间，记录时间，技术亮点，项目图片
- 测点粗视图
  - 测点id，测点名
- 测点细视图
  - 测点id，测点名，时间，温度，湿度，能耗，PM2.5, CO2，室外温度，室内温度，项目名
- 设备视图
  - 设备id，测点名，设备名，时间，测点相关数据...

## 常用数据操作

网站所允许的数据操作由封装的api提供。要求api提供比网站所需更多的操作。

- 网站操作
  - 以项目名查询项目粗略视图
  - 以项目名查询项目细视图
  - 以项目id查询项目粗略视图
  - 以项目id查询项目细视图

- 以测点查询测点视图
- 以测点名查询测点视图
- 以设备id查询设备视图
- 以测点名查询设备视图
- api封装操作
  - 所有上述的网站操作
  - 数据库更新操作
  - 数据库插入操作

## 优化查询速度方案

历史记录很少会需要更新或修改，所以所有的设备历史记录，测点历史记录，项目历史记录都可以添加index来加速查询操作。

## 问题

- 测点是不是以视图方式组织取决于建研院的基本数据是由子设备提供还是子设备数据的总结。

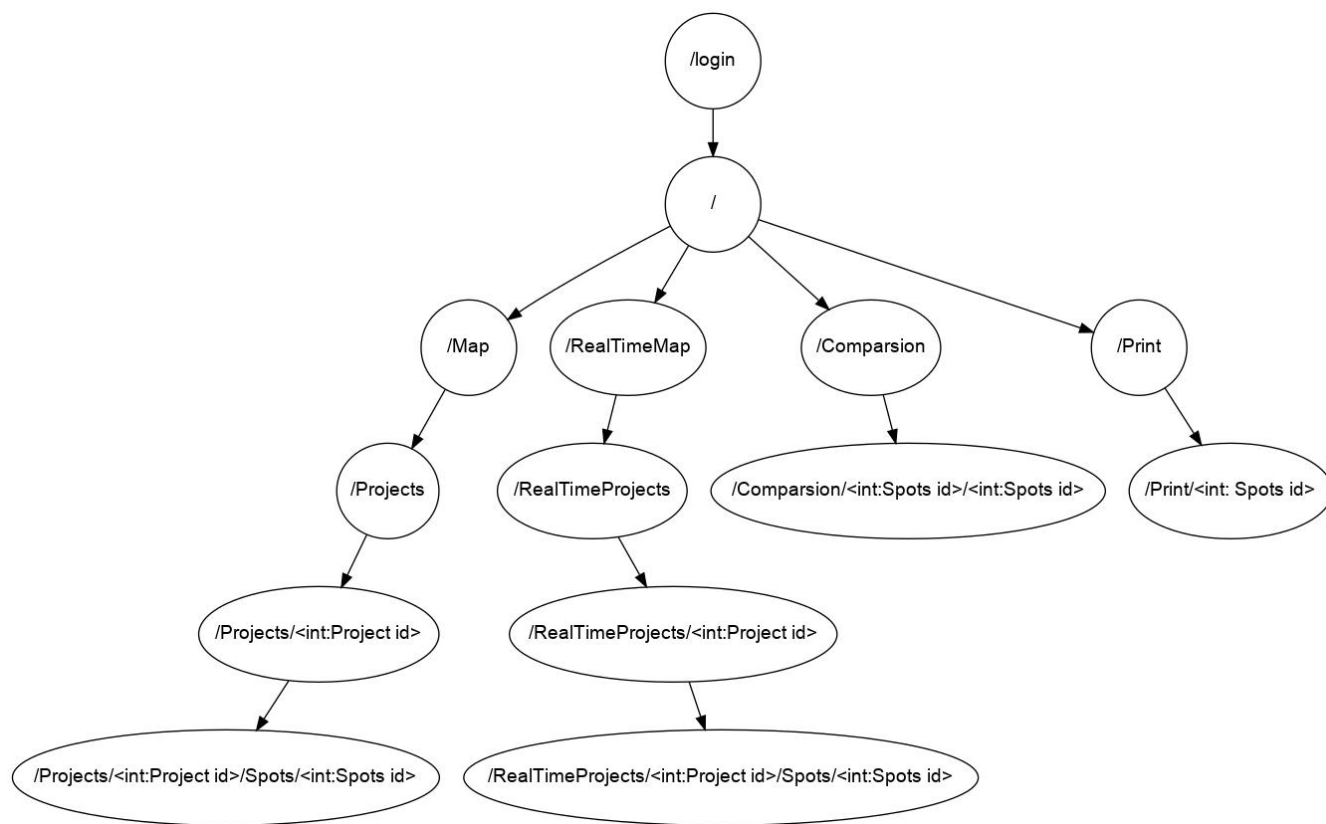
# 十三五长江流域建筑供暖解决方案展示平台后台设计

## 后台目标

- 封装数据库常用视图查询
- 提供数据插入和更新接口（每七天更新一次）
- 实时数据缓存和转发
- 登陆功能
- 网站routing
- 网页rendering
- http错误页面处理

## 后台url mapping

- url映射方案



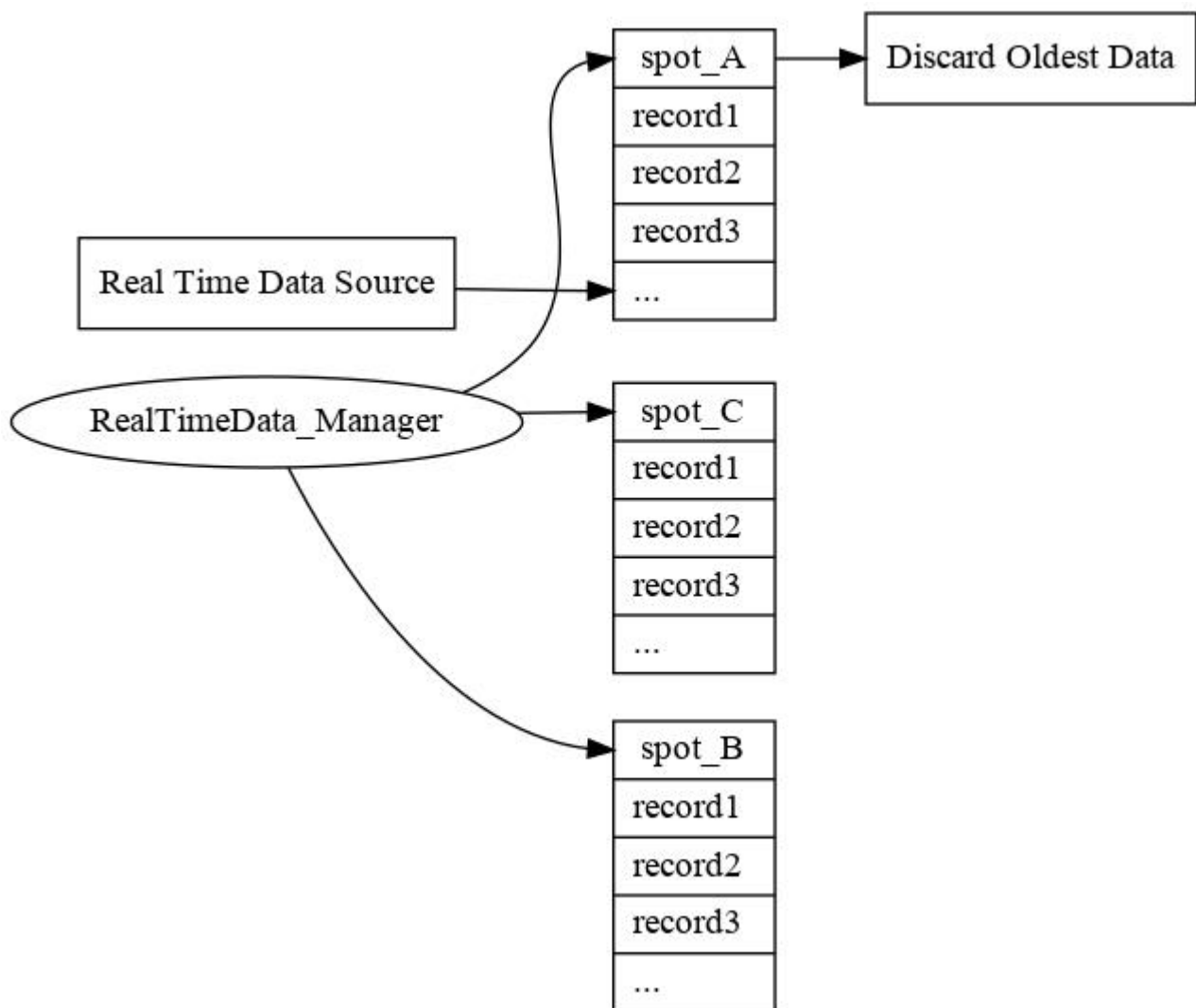
#### url和数据库视图的对应

- /login -> NULL
- / -> (待定)
- /Map -> 项目粗视图
- /RealTimeMap -> (待定, 实时数据)
- /Comparsion -> NULL
- /Print -> NULL
- /Projects -> 项目粗视图
- /Projects/<int:Project id> -> 测点粗视图, 项目细视图
- /Projects/<int:Project id>/Spots/<int:Spots id> -> 测点细视图
- /RealTimeProjects -> (待定, 实时数据)
- /RealTimeProjects/<int:Project id> -> (待定, 实时数据)
- /RealTimeProjects/<int:Project id>/Spots/<int:Spots id> -> (待定, 实时数据)
- /Comparsion/<int:Spots id>/<int:Spots id> -> 测点细视图 x2
- /Print/<int: Spots id> -> 测点细视图

#### 实时数据模块

实时数据将从支持数据实时查询的设备收集。实时数据模块需要实现数据收集接口, 负责从实时数据源每五分钟收集一次数据。单个测点的实时数据在后台用queue的形式维护, queue的最大长度为100, 每五分钟向queue内添加一条实时记录。如果queue已满, 将最早的记录弹出, 再将新记录加入队尾。通过用queue作缓冲区的方式, 实时数据模块可以实现在任何一个时段查询单个实时设备都能获得从那个时间点向前100条的数据, 方便前端查询和展示。每个实时测点都有自己单独的queue, 所有测点由一个<实时测点列表>管理。

除了收集和管理实时数据, 实时数据模块还需要提供类似数据库视图的数据查询api, 后台可以通过调用该api以测点名或测点id查询到测点的实时视图。



实时测点列表由RealTimeDataManager内统一管理。RealTimeDataManager向下每五分钟发送一次queue更新请求，向上每接收到一次数据展示请求根据请求的测点id返回对应的测点queue内数据。若是第一次请求便返回所有数据，若不是则返回最新获取的数据。

#### 数据库数据量统计

- 预计单条测点数据大小为116字节

Field	Size/bytes
spot_id	8
project_id	8
time	24
name	20
tempreture	8
humidity	8
energy consumption	8
PM2.5	8
CO2	8
outdoor tempreture	8
outdoor humidity	8

Field	Size/bytes
TOTAL	116

- 每年数据4500万条，每年测点总数据为  $\frac{116 \times 45,000,000}{1024^3} = 4.9GB$
- 相比于测点的数据量级，项目信息和用户信息所占空间可以忽略不记。
- 支持实时数据的测点96个（环科智控），单个测点内存里同时会存储100条记录，一共  $\frac{96 \times 116 \times 100}{1024^2} = 2.1MB$
- 实时数据更新时数据流量21KB

## 缺失数据处理

- 两套数据，数据库存储真实数据，展示前端显示展示数据
- 数据缺失情况
  - 单条记录缺失列
    - 有温湿度，无电耗
      - 联想根据室外温度和温湿度插入预测电耗
    - 有电耗，无温湿度
      - 根据室外温度和电耗插入预测温湿度
  - 时间段内缺失整条记录
    - 缺失3小时以上不用补充
    - 缺失三小时以下：
      - 若查询精确度到日
        - 缺失数据10条以内
          - 根据前后时段的数据设置合理上下限，插入上下限内的随机数据（或直接连线）。
        - 缺失数据10条以上
          - 暂无数据
      - 若查询精确度到月或年
        - 首先检查是否每天都是数据（取当日平均）
          - 若某天没有数据，根据前后两天的平均数据插入预测数据。