## 变更记录

|  |  |  |
| --- | --- | --- |
| 变更者 | 变更日期 | 变更内容 |
| 王涛 | 2014-04-24 | 创建 |
| 王涛 | 2014-06-13 | 修改，新增对各表的文字描述  weather\_f\_lifeindex表增加lifeindex\_name字段 |
| 王涛 | 2014-06-19 | 增加表atmosphere\_aqi |
| 王涛 | 2014-06-20 | 修改表weather\_h，rainfall的类型改为double |
| 王涛 | 2014-08-16 | 删除表lifeindex，修改表weather\_f\_lifeindex |
| 王涛 | 2014-08-25 | 新增表disaster\_alarm |
| 王涛 | 2014-09-18 | 修改表atmosphere\_aqi,新增pm25字段 |
| 王涛 | 2014-09-18 | 新增表atmosphere\_aqi\_d |

## STRUC项目概要

**已采集情况概述：**

天气小时(包括实时)，来自[国家气象局](http://weather.com.cn/)。

天气预报(包括生活指数)，来自[国家气象局](http://weather.com.cn/)。

全国空气质量，包括最近24小时，最近30天的全国289个城市的空气质量数据，数据粒度为站点，来自[PM2.5监测网](http://www.cnpm25.cn/)。

浙江废气废水，来自[浙江省自行监测平台](http://app.zjepb.gov.cn:8089/nbjcsj/)。

杭州交通数据，数据每5分钟发布一次，来自[杭州市交通拥堵指数实时监测平台](http://www.hzjtydzs.com/web/current.aspx)。

浙江空气质量，来自[浙江省环保厅webservice](http://218.108.43.220:9090/services/QueryDataWebServiceUtil/queryCityAQIReport?beginDate=20131127&endDate=20131128&code=330100,330200,330300,330400,330500,330600,330700,330800,330900,331000,331100)，包含天和小时数据，数据粒度包括城市、区县、站点。

地表水饮用水，来自衢州Webservice接口，数据每4小时发布，粒度为站点。

全国灾害预警，来自[国家气象局](http://weather.com.cn/)。

**计划采集情况概述：**

全国空气质量,来自[PM25.in](http://pm25.in/)。

天气历史，来自[2345天气网](http://tianqi.2345.com/)。

## MySQL表设计

### 行政区city

全国县及县级以上行政区划数据

未来可能把粒度细化到乡镇级别

来自：[国家统计局](http://www.stats.gov.cn/tjsj/tjbz/xzqhdm/201401/t20140116_501070.html)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| chief\_code | varchar 12 | Y | 行政区编码 |  |
| name | varchar 20 | Y | 行政区名称 |  |
| pinyin | varchar 60 | Y | 行政区名称的拼音 |  |
| jianpin | varchar 20 | Y | 行政区名称的简拼 |  |
| area\_code | varchar 5 | Y | 区号 |  |
| post\_code | char 6 | Y | 邮政编码 |  |
| level | tinyint | Y | 行政区级别  0代表国家  1代表一级行政区  2代表二级行政区  3代表三级行政区  4代表四级行政区  以此类推 |  |
| pid | int | Y | 上级行政区编码 |  |

### 天气-小时weather\_h

全国天气小时数据

来自：[国家气象局](http://weather.com.cn/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| temperature | smallint | Y | 温度 |  |
| humidity | smallint | Y | 湿度 |  |
| wind\_force | tinyint | Y | 风力,数字代表等级 |  |
| wind\_direction | smallint | Y | 风向 |  |
| rainfall | double | Y | 降雨量 |  |
| publish\_time | datetime | Y | 发布时间，代表当前记录对应时间点 | 如2014-06-12 11:00:00 |
| collect\_time | datetime | Y | 采集时间 |  |
| city\_id | int | Y | 行政区编码，无符号 | 外键 |

### 天气-预测 weather\_f

全国天气预报数据，包括逐6小时天气预报和七天天气预报

网站在每天都8:00 11:00 18:00发布数据，爬虫会在半个小时后采集一次数据

来自：[国家气象局](http://weather.com.cn/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| date | varchar 10 | Y | 预测日期（dd号） | 12号 |
| datetime | date | Y | 预测日期（yyyy-MM-dd） | 2014-06-18 |
| time\_period | varchar 30 | Y | 时间段 | 白天/夜间/ 12日14:00-20:00 |
| phenomena | varchar 20 | Y | 天气现象 |  |
| temperature | varchar 20 | Y | 温度 |  |
| wind\_force | varchar 20 | Y | 风力 |  |
| wind\_direction | varchar 20 | Y | 风向 |  |
| type | tinyint | Y | 0代表逐6小时天气预报  1代表七天天气预报 |  |
| publish\_time | datetime | Y | 发布时间，天气预报数据的发布时间 | 每天8:00  11:00  18:00发布 |
| collect\_time | datetime | Y | 采集时间 |  |
| city\_id | int | Y | 行政区编码，无符号 | 外键 |

### 生活指数表 weather\_f\_lifeindex

全国天气『当日生活指数』数据,包括逐9个热门生活指数和其他11个生活指数

数据同天气预测数据除同一张网页

如网站新增了新的生活指数，程序中会相应的增加）

来自：[国家气象局](http://weather.com.cn/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| publish\_time | datetime | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| city\_id | int | Y | 行政区编码，无符号 | 外键 |
| name | varchar 20 | Y | 名称，如“太阳镜指数” |  |
| value | varchar 10 | Y | 值，如“需要” |  |
| description | varchar 100 | Y | 描述，如“白天根据户外光线情况佩戴太阳镜” |  |

### 浙江空气监测点zj\_atmosphere\_monitorsite

浙江省空气数据，粒度包括站点级、城市级、区县级

来自：[浙江省环保厅](http://218.108.43.220:8099/aqi/services/listServices)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| site\_code | char 16 | Y | 站点编码 |  |
| name | varchar 30 | Y | 名称 |  |
| area\_type | tinyint | Y | 区域类型  1=站点  2=城市  3=区县 |  |
| city\_id | int | Y | 所在地编码 | 外键 |

### 浙江空气质量 zj\_atmosphere

浙江省空气数据，包含每小时的和每天的（最新获得当前时间前一天）

由于数据源自省厅大气webservice接口数据，但此接口无法通过外网访问，需通过衢州webservice中转一下

来自：浙江省环保厅

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| so2\_data | double | Y | so2浓度 微克/立方 |  |
| so2\_iaqi | double | Y | so2 IAQI 指数 |  |
| no2\_data | double | Y | no2浓度 微克/立方 |  |
| no2\_iaqi | double | Y | no2 IAQI指数 |  |
| pm10\_data | double | Y | pm10浓度 微克/立方 |  |
| pm10\_iaqi | double | Y | pm10 IAQI指数 |  |
| co\_data | double | Y | co浓度 毫克/立方 |  |
| co\_iaqi | double | Y | co IAQI指数 |  |
| o3\_1\_data | double | Y | o3 1小时浓度 微克/立方 |  |
| o3\_1\_iaqi | double | Y | o3 1小时IAQI指数 |  |
| o3\_8\_data | double | Y | o3 8小时浓度 微克/立方 |  |
| o3\_8\_iaqi | double | Y | o3 8小时IAQI指数 |  |
| pm25\_data | double | Y | pm2.5 浓度 微克/立方 |  |
| pm25\_iaqi | double | Y | pm2.5 IAQI指数 |  |
| aqi\_value | double | Y | AQI指数,所有因子中IAQI最大值 |  |
| quality\_status | tinyint | Y | 0代表优  1代表良  2代表轻度污染  3代表中度污染  4代表重度污染  5代表严重污染 |  |
| quality\_level | tinyint | Y | 0代表一级  1代表二级  2代表三级  3代表四级  4代表五级  5代表六级 |  |
| main\_ pollutant | varchar 20 | Y | 主要污染物 |  |
| type | tinyint | Y | 0代表天  1代表小时 |  |
| topn | smallint | Y | 排名 |  |
| publish\_time | varchar 30 | Y | 发布时间，代表当前记录对应时间点 | 天数据—>2014-06-12  小时数据—>2014-05-23,13 |
| collect\_time | datetime | Y | 采集时间 |  |
| zj\_atmosphere\_monitorsite\_id | int | Y | 浙江空气监测站点编码 | 外键 |

### 空气质量atmosphere

全国空气质量数据，粒度是监测站点级的，包括最近24小时，最近30天的数据

来自：[PM2.5监测网](http://www.cnpm25.cn/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 |  |
| aqi | int | Y | AQI指数,所有因子中IAQI最大值 | 主键 |
| so2 | double | Y | so2浓度 微克/立方米 |  |
| no2 | double | Y | no2浓度 微克/立方米 |  |
| pm10 | double | Y | pm10浓度 微克/立方米 |  |
| co | double | Y | co浓度 毫克/立方米 |  |
| o3 | double | Y | o3浓度 微克/立方米 |  |
| pm25 | double | Y | pm25浓度微克/立方米 |  |
| topn | smallint | Y | 排名 |  |
| main\_ pollutant | varchar10 | Y | 主要污染物 |  |
| type | tinyint | Y | 0代表天  1代表小时 |  |
| publish\_time | varchar 30 | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| atmosphere\_monitorsite\_id | int | Y | 空气监测站点编码 | 外键 |

### 空气监测站点atmosphere\_monitorsite

城市为地级市级别，其下包括若干空气监测站点，城市289个，监测站点941个

来自：[PM2.5监测网](http://www.cnpm25.cn/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| name | varchar 30 | Y | 名称 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 全国城市AQI表 atmosphere\_aqi

这边的『城市』包括全国的主要直辖市、地级市、县级市。采集时aqi、pm2.5时均值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| aqi | int | Y | aqi值，整数 |  |
| pm25 | int | Y | PM2.5浓度,单位μg/m³ |  |
| publish\_time | datetime | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 全国城市AQI日均值表 atmosphere\_aqi\_d

这边的『城市』包括全国的主要直辖市、地级市、县级市。采集aqi日均值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| aqi | int | Y | aqi值，整数 |  |
| publish\_time | date | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 路 road

杭州城区各路段数据

双向共226条路

来自[杭州市交通拥堵指数实时监测平台](http://www.hzjtydzs.com/web/current.aspx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| name | varchar 100 | Y | 路名 |  |
| level | tinyint | Y | 道路等级 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 交通traffic

杭州交通拥堵数据，粒度：路

来自：[杭州市交通拥堵指数实时监测平台](http://www.hzjtydzs.com/web/current.aspx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| jam\_rank | tinyint | Y | 拥堵等级 |  |
| current\_speed | double | Y | 当前速度 |  |
| last\_hour\_avg\_speed | double | Y | 上小时均速 |  |
| last\_week\_avg\_speed | double | Y | 上周均速 |  |
| publish\_time | datetime | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| road\_id | int | Y | 路的编码 | 外键 |

### 衢州水质监测站点water\_monitorsite

衢州地表水、饮用水监测站点的信息

来自：衢州WebService接口

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| name | varchar 30 | Y | 名称 |  |
| address | varchar 100 | Y | 地址 |  |
| junction\_section | varchar 30 | Y | 交界断面 |  |
| type | tinyint | Y | 类型  0地表水  1饮用水 |  |
| longitude | double | Y | 经度 |  |
| latitude | double | Y | 纬度 |  |
| quality\_level | tinyint | Y | 水质等级 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 地表水surface\_water

衢州地表水数据，每4个小时产生一个记录，即每日的0:00 4:00 8:00 12:00 16:00 20:00会发布一个数据记录

来自：衢州WebService接口

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| ph | double | Y | ph值 |  |
| dissolved\_oxygen | double | Y | 溶解氧 |  |
| ammonia\_nitrogen | double | Y | 氨氮 |  |
| permanganate | double | Y | 高锰酸盐 |  |
| turbidity | double | Y | 浊度 |  |
| temperature | double | Y | 水温 |  |
| conductivity | double | Y | 电导率 |  |
| phosphorus | double | Y | 总磷 |  |
| nitrogen | double | Y | 总氮 |  |
| water\_level | double | Y | 水位 |  |
| flow | double | Y | 流量 |  |
| main\_pollutant | varchar 20 | Y | 主要污染物 |  |
| publish\_time | datetime | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| water\_monitorsite\_id | int | Y | 水质监测站点编码 | 外键 |

### 饮用水drink\_water

衢州地表水数据，每4个小时产生一个记录，即每日的0:00 4:00 8:00 12:00 16:00 20:00会发布一个数据记录

来自：衢州WebService接口

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| ph | double | Y | ph值 |  |
| dissolved\_oxygen | double | Y | 溶解氧 |  |
| ammonia\_nitrogen | double | Y | 氨氮 |  |
| permanganate | double | Y | 高锰酸盐 |  |
| phosphorus | double | Y | 总磷 |  |
| nitrogen | double | Y | 总氮 |  |
| flow | double | Y | 流量 |  |
| fluoride | double | Y | 氟化物 |  |
| cod | double | Y | 化学需氧量 |  |
| hexavalent\_chromium | double | Y | 六价铬 |  |
| chlorophyll | double | Y | 叶绿素 |  |
| publish\_time | datetime | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| water\_monitorsite\_id | int | Y | 水质监测站点编码 | 外键 |

### 浙江废气监测站zj\_pollution\_gas\_monitorsite

当前站点数297

来自[浙江省企业自行监测平台](http://app.zjepb.gov.cn:8089/nbjcsj/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 站点编码，自增 | 主键 |
| name | varchar100 | Y | 名称 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 浙江废水监测站zj\_pollution\_water\_monitorsite

当前站点数478

来自[浙江省企业自行监测平台](http://app.zjepb.gov.cn:8089/nbjcsj/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 站点编码，自增 | 主键 |
| name | varchar100 | Y | 名称 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

### 浙江废气zj\_pollution\_gas

浙江省企业自行监测废气站点数据，分为小时数据和天数据

来自[浙江省企业自行监测平台](http://app.zjepb.gov.cn:8089/nbjcsj/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| nox | double | Y | 氮氧化物 毫克/立方 |  |
| fume | double | Y | 烟尘 毫克/立方 |  |
| so2 | double | Y | 二氧化硫 毫克/立方 |  |
| type | tinyint | Y | 0代表天  1代表小时 |  |
| publish\_time | varchar 30 | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| zj\_pollution\_gas\_monitorsite\_id | int | Y | 浙江废气监测站编码 |  |

### 浙江废水zj\_pollution\_water

浙江省企业自行监测废水站点数据，分为小时数据和天数据

来自[浙江省企业自行监测平台](http://app.zjepb.gov.cn:8089/nbjcsj/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| ph | double | Y | PH值 |  |
| cod | double | Y | 化学需氧量 毫克/升 |  |
| an | double | Y | 氨氮 毫克/升 |  |
| type | tinyint | Y | 0代表天  1代表小时 |  |
| publish\_time | varchar 30 | Y | 发布时间 |  |
| collect\_time | datetime | Y | 采集时间 |  |
| zj\_pollution\_water\_monitorsite\_id | int | Y | 浙江废水监测站编码 |  |

### 全国灾害预警disaster\_alarm

全国灾害预警

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| type | tinyint | Y | 预警种类  1代表台风  2代表暴雨  3代表暴雪  4代表寒潮  5代表大风  6代表沙尘暴  7代表高温  8代表干旱  9代表雷电  10代表冰雹  11代表霜冻  12代表大雾  13代表霾 |  |
| level | tinyint | Y | 预警等级  1代表蓝色  2代表黄色  3代表橙色  4代表红色 |  |
| title | varchar200 | Y | 标题 |  |
| description | text | Y | 描述内容 |  |
| publish\_time | datetime | Y | 发布时间 |  |
| collect\_time | datetime | N | 采集时间 |  |
| city\_id | int | Y | 所属行政区编码 | 外键 |

## 计划新增

### 天气-天weather\_d

『暂未采集』

全国天气数据，粒度是天，数据范围从2011年1月1号到当前时间的前一天

来自：[2345天气网](http://tianqi.2345.com/)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| phenomena | varchar 20 | Y | 天气现象 |  |
| max\_temperature | tinyint | Y | 最高温度 -128~127℃ |  |
| min\_temperature | tinyint | Y | 最低温度 -128~127℃ |  |
| wind\_force | varchar 20 | Y | 风力 |  |
| wind\_direction | varchar 20 | Y | 风向 |  |
| publish\_time | date | Y | 发布时间，代表当前记录对应时间点 | 如2014-06-12 |
| collect\_time | datetime | Y | 采集时间 |  |
| city\_id | int | Y | 行政区编码，无符号 | 外键 |

### 空气质量表atmosphere\_in

『暂未采集』

来自<http://www.pm25.in/>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名称 | 字段类型 | 允许为空 | 说明 | 备注 |
| id | int | N | 无含义，自增 | 主键 |
| aqi | smallint | Y | 空气质量指数(AQI)，即air quality index，是定量描述空气质量状况的无纲量指数 |  |
| area | varchar 10 | Y | 城市名称 |  |
| position\_name | varchar 10 | Y | 监测点名称 |  |
| station\_code | char 5 | Y | 监测点编码 |  |
| so2 | smallint | Y | 二氧化硫1小时平均 |  |
| so2\_24h | smallint | Y | 二氧化硫24小时滑动平均 |  |
| no2 | smallint | Y | 二氧化氮1小时平均 |  |
| no2\_24h | smallint | Y | 二氧化氮24小时滑动平均 |  |
| pm10 | smallint | Y | 颗粒物（粒径小于等于10μm）1小时平均 |  |
| pm10\_24h | smallint | Y | 颗粒物（粒径小于等于10μm）24小时滑动平均 |  |
| co | double | Y | 一氧化碳1小时平均 |  |
| co\_24h | double | Y | 一氧化碳24小时滑动平均 |  |
| o3 | smallint | Y | 臭氧1小时平均 |  |
| o3\_24h | smallint | Y | 臭氧日最大1小时平均 |  |
| o3\_8h | smallint | Y | 臭氧8小时滑动平均 |  |
| o3\_8h\_24h | smallint | Y | 臭氧日最大8小时滑动平均 |  |
| pm2\_5 | smallint | Y | 颗粒物（粒径小于等于2.5μm）1小时平均 |  |
| pm2\_5\_24h | smallint | Y | 颗粒物（粒径小于等于2.5μm）24小时滑动平均 |  |
| primary\_pollutant | varchar 30 | Y | 首要污染物 |  |
| quality | tinyint | Y | 空气质量指数类别，有“优、良、轻度污染、中度污染、重度污染、严重污染”6类 |  |
| time\_point | datetime | Y | 数据发布的时间 |  |