

本工程为海气相关数据爬取的汇总项目

```
.  
├── main.py //  
└── ioc_station_spider.py // 爬取 ioc station 数据爬虫
```

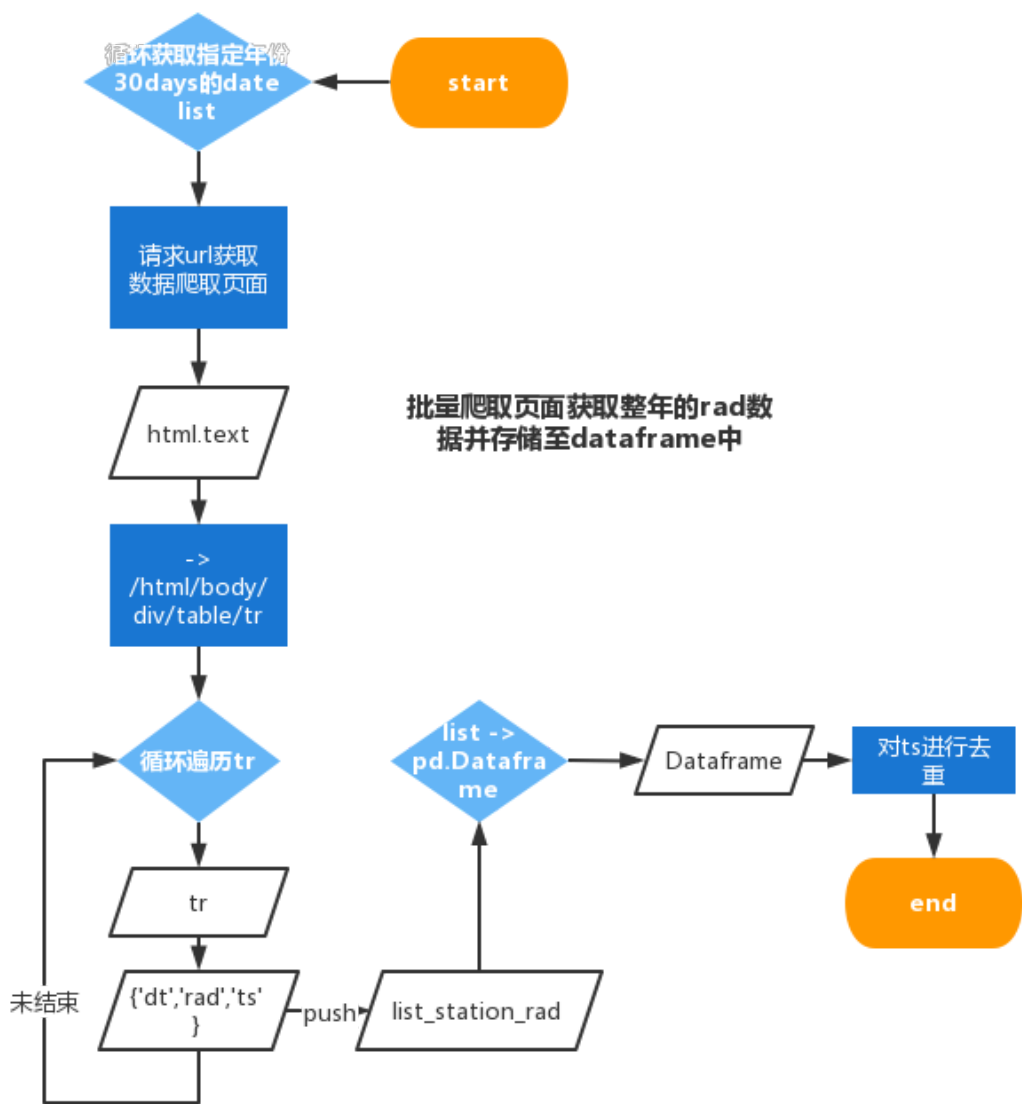
1- `ioc_station_spider` - 爬取 ioc station 数据爬虫

爬取ioc 海洋站rad数据，爬取地址为: <https://www.ioc-sealevelmonitoring.org/bgraph.php?code=shek&output=tab&period=30&endtime=2022-01-01>

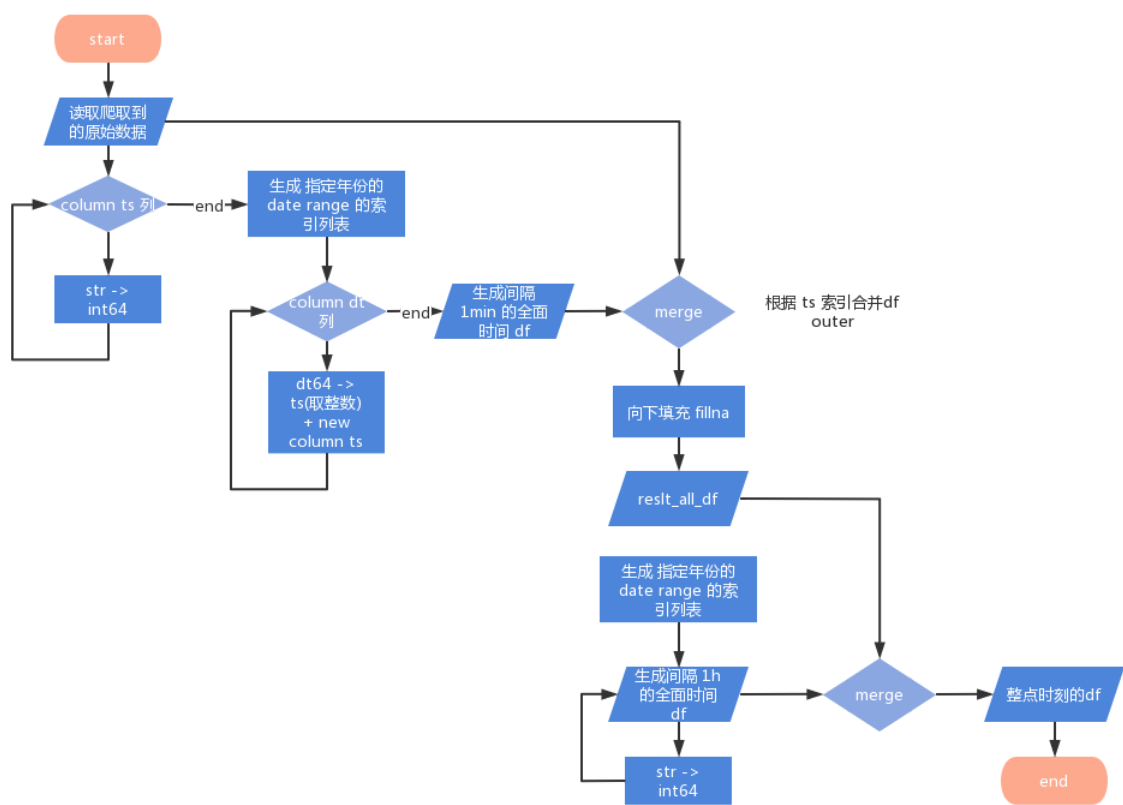
```
options:  
  code: // 站名  
  period: // 以endtime向前推 x 天  
  endtime: // 结束时间(utc)
```

以下为主要步骤:

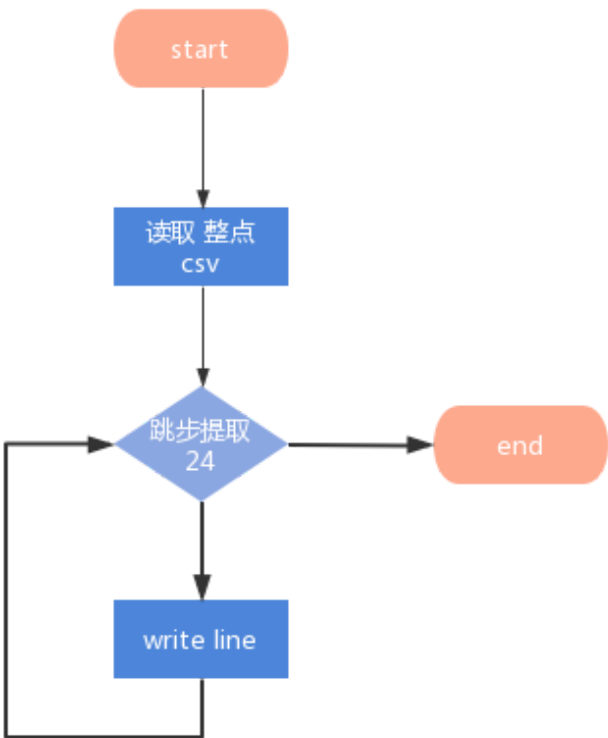
step-1: ioc网站爬取数据



step-2: 数据标准化



step-3: csv转txt



生成的结果集:
文件格式

```
index //标记行
dt //对应时间(utc)
ts //时间戳
rad // 爬取的海洋站的观测值
```

目录结构

```
.
├── qinglan_2021_source // 样例存储子目录
│   ├── qinglan_2021_all.csv // 爬取的原始文件
│   ├── qinglan_2021_fill_minutes.csv // fillna 填充后的指定年份的 sep为1min 数据集
│   ├── qinglan_2021_split_hours.csv // 提取 sep为1h的数据集
│   └── qinglan_2021_split_hours.txt // sep=1h -> txt 每行对应当日的24个时刻的数据
```