

## 甲方名称:

某国企 (因保密协议限制无法公开具体信息)

## 甲方需求:

根据需求定制一采集平台, 并进行长期维护。

1. 采集目标: 携程航班动态、非常准航班动态;
  2. 从数据库读取待查询航线(出发城市-到达城市), 分别在各网站进行搜索, 获取每条航线的执飞航班的航班号等信息, 再结合查询日期, 查询获取每个航班的航班动态数据;
  3. 提取搜索结果的"航班信息","航班时间","航班状态","前程航班信息"存储在数据库中;
  4. 采集频率: 每日启动一次采集任务, 24小时不间断查询采集, 保留每批次采集结果;
  5. 采集量: 每天采集航班动态数据超100万条;
- 航班动态信息属性如下JSON文档所示:

```
{  
  
  "destTerminal": "T2B",  
  
  "deptGate": "11A",  
  
  "aDeptDate": "2017-04-17",  
  
  "destDate": "2017-04-17",  
  
  "preDest": "青岛",  
  
  "eDestDate": "2017-04-17",  
  
  "preDept": "南昌",  
  
  "airlineCode": "ZH",  
  
  "preFlightStatus": "计划",  
  
  "eDeptDate": "2017-04-17",  
  
  "punctualityRate": "77%",  
  
  "planeType": "738",  
  
  "preDeptDelay": 0,  
  
  "flyTime": 2.75,  
  
  "preFlightNo": "SC4910",  
  
  "destAirport": "江北国际机场",  
  
  "destDelayTime": 0,
```

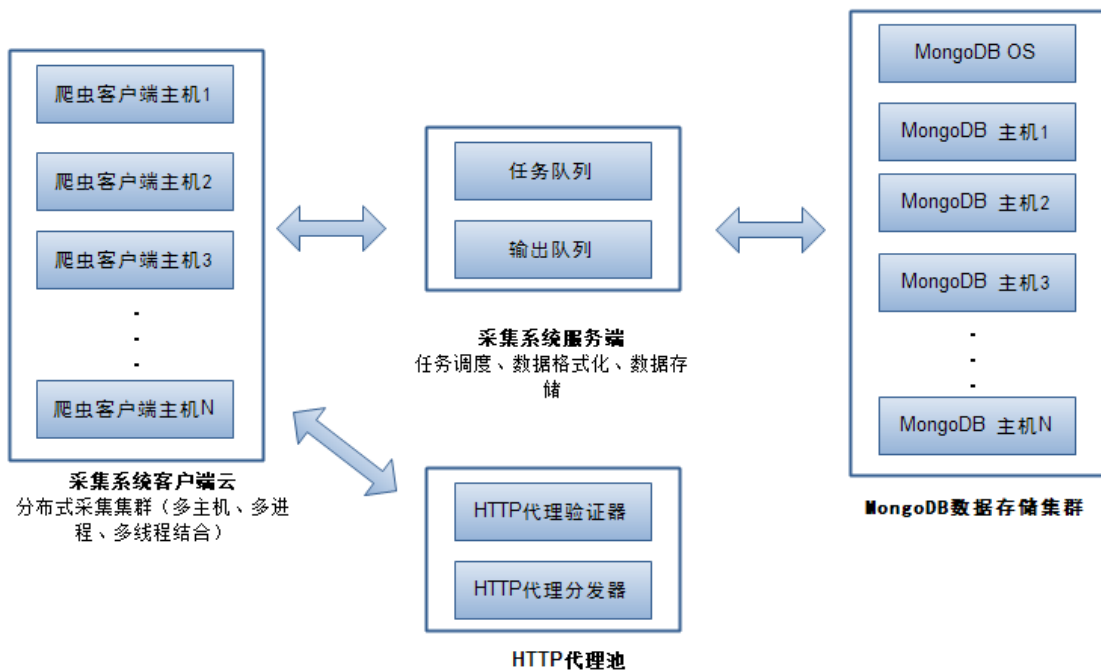
```
"carousel": "--",  
  
"destAptCode": "CKG",  
  
"ckiCounter": "A",  
  
"deptDate": "2017-04-17",  
  
"flightStatus": "计划",  
  
"dest": "重庆",  
  
"deptAptCode": "TAO",  
  
"deptDelayTime": 0,  
  
"deptCode": "TAO",  
  
"destCode": "CKG",  
  
"std": "19:00",  
  
"atd": "",  
  
"sta": "21:45",  
  
"ata": "",  
  
"flyDistance": 1577,  
  
"airlineName": "深圳航空",  
  
"deptAirport": "流亭国际机场",  
  
"deptTerminal": "T1",  
  
"preDestDelay": 0,  
  
"dept": "青岛",  
  
"eta": "21:45",  
  
"etd": "19:00",  
  
"aDestDate": "2017-04-17",  
  
"flightNo": "ZH3027"  
  
}
```

### 技术难点：

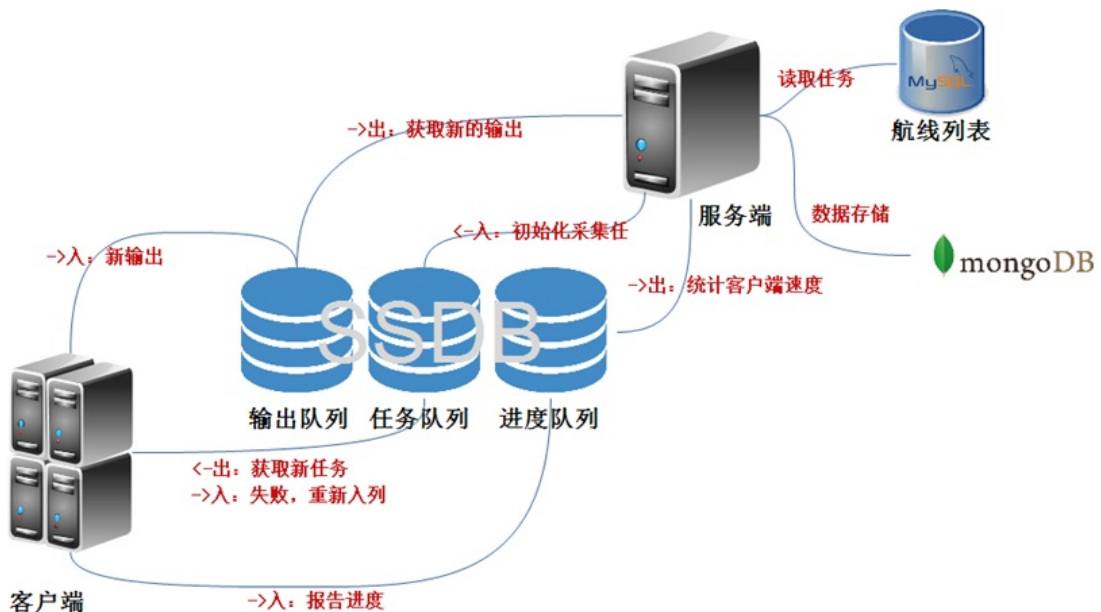
- 查询任务量大，每天查询航线超过5000条，每条航线涉及数个到数十个航班；
- 每批次的查询包括T天（查询当日）和T+1天（查询日的次日）的查询任务；
- 要支持订阅航班的航班动态查询（即随时指定一条新的航线，此后将作为常规航线的查询任务）；
- 24小时不间断地查询，直至该航班完成所有航程之后才结束，所以数据输出量比较大，每天输出记录数量超过100W条。

### 实现方案：

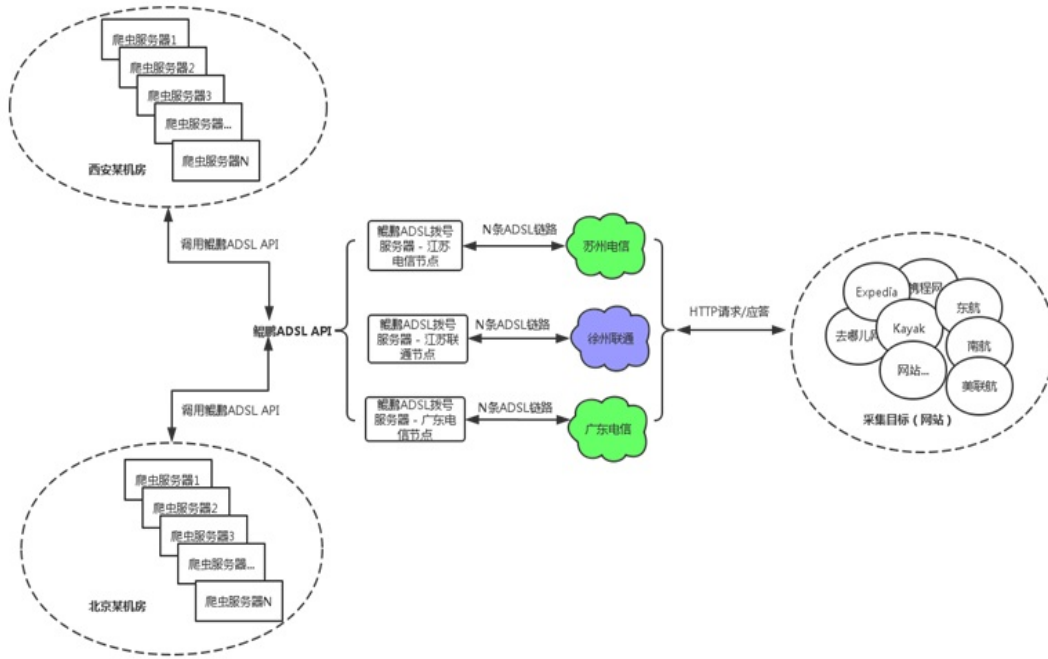
- 通过研究网站的防采集策略，结合鲲鹏ADSL动态IP代理可以有效并稳定持续地采集这些网站的数据；
- 通过Webkit技术有效绕过各种加密策略；
- 采用分布式架构，可以动态扩展采集集群规模（增加采集机器）；
- 采用ActiveMQ作为消息队列，采用MongoDB作为输出数据库；



系统架构图



## 系统运行流程图



鲲鹏ADSL动态IP代理系统原理图

### 项目状态:

本项目于2015年7月1日正式开始上线运行，目前仍在稳定进行中。



**西安鲲鹏之鹏网络信息技术有限公司**

**选择我们，所有数据都是你的！**



公司名称: 西安鲲鹏之鹏网络信息技术有限公司

网 址: <http://www.site-digger.com/>

地 址: 陕西省西安市雁塔区长安中路99号长安文化综合大厦11821室

邮 编: 710061

联系电话: 029 - 87553281

手 机: 13571845363 齐工

13389148466 周工

客 服 QQ: 1649677458 或 312602670

邮 箱: hello@site-digger.com