Python日志处理

woniu@Reboot
@知乎
@github

- 怎么统计ip出现次数的前10?

- 数据统计

- 入库

- 展现

- 如何做数据统计

  - 读取日志

  - 累加统计

  - 打印前10

日志

- 61.159.140.123 - - [23/Aug/2014:00:01:42 +0800] "GET /favicon.ico HTTP/1.1" 404 \ "-" "Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.66 Safari/537.36 LBBROWSER" "-"

- 61.159.140.123 - - [23/Aug/2014:00:01:42 +0800] "GET /favicon.ico HTTP/1.1" 404 \ "-" "Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.66 Safari/537.36 LBBROWSER" "-"

- 61.159.140.123 - - [23/Aug/2014:00:01:42 +0800] "GET /favicon.ico HTTP/1.1" 404 \ "-" "Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.66 Safari/537.36 LBBROWSER" "-"

- 61.159.140.123 - - [23/Aug/2014:00:01:42 +0800] "GET /favicon.ico HTTP/1.1" 404 \ "-" "Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/29.0.1547.66 Safari/537.36 LBBROWSER" "-"

- 66.249.64.5 - - [23/Aug/2014:00:02:16 +0800] "GET /data/uploads/2013/0519/09/small_51982ba18e012.jpg HTTP/1.1" 200 \ "-" "Googlebot-Image/1.0" "-"

- 66.249.64.10 - - [23/Aug/2014:00:02:54 +0800] "GET /data/uploads/2013/0319/08/middle_5147b116e93b4.jpg HTTP/1.1" 200 \ "-" "Googlebot-Image/1.0" "-"

统计

- 解析每一行

- 用dict做数据统计

  - 出现一次+1

- 变成list，排序

- 打印前10

```python
f = open('www_access_20140823.log')
res = {}
for l in f:
    arr = l.split(' ')
    ip = arr[0]
    res[ip] = res.get(ip,0)+1
res_list = res.items()

for k in sorted(res_list,key=lambda x:x[3],reverse=True)[:10]:
    print k
```

多维度统计

- 按照IP，status和url三个维度统计

```python
f = open('../../www_access_20140823.log')
res = {}
for l in f:
    arr = l.split(' ')
    ip = arr[0]
    url = arr[6]
    status = arr[8]
    res[(ip,url,status)] = res.get((ip,url,status),0)+1
res_list = [(k[0],k[1],k[2],v) for k,v in res.items()]

for k in sorted(res_list,key=lambda x:x[3],reverse=True)[:10]:
    print k
```

```
('222.86.153.12', '/images/cursor_minify.cur', '404', 60)
('222.86.153.12', '/images/cursor_zoom.cur', '404', 32)
('58.253.6.133', '/images/cursor_minify.cur', '404', 32)
('111.85.34.165', '/%3Ca%20href=', '404', 28)
('58.253.6.133', '/images/cursor_zoom.cur', '404', 27)
('218.29.111.117', '/images/cursor_zoom.cur', '404', 27)
('218.29.111.117', '/images/cursor_minify.cur', '404', 26)
('117.63.146.40', '/public/js/common.js?20110824', '200', 19)
('117.63.146.40', '/favicon.ico', '404', 18)
('117.63.146.40', '/public/js/weibo.js?20110824', '200', 16)
[Finished in 0.3s]
```

web端显示

- 入库

- web显示

入库

```
-- 统计log
drop table if exists log;
create table log(
    ip varchar(400),
    url varchar(200),
    status int,
    value int
);


  for s in res_list:
      # if s[3]<5:
          # continue
      sql = 'insert log values ("%s","%s",%s,%s)' % s
      print sql
      try:
          db.execute(sql)
      except Exception, e:
          pass
```

简易的web显示

```python
@app.route('/table')
def table():
    table = '<table border="1">'
    c = db.execute('select * from log order by value desc limit 20;')
    for l in c.fetchall():
        table += '<tr><td>%s</td><td>%s</td><td>%s</td><td>%s</td></tr>'%l
    table +='</table>'
    return table
```

222.86.153.12	/images/cursor_minify.cur	404	60
222.86.153.12	/images/cursor_zoom.cur	404	32
58.253.6.133	/images/cursor_minify.cur	404	32
111.85.34.165	/%3Ca%20href=	404	28
58.253.6.133	/images/cursor_zoom.cur	404	27
218.29.111.117	/images/cursor_zoom.cur	404	27
218.29.111.117	/images/cursor_minify.cur	404	26
117.63.146.40	/public/js/common.js?20110824	200	19
117.63.146.40	/favicon.ico	404	18
117.63.146.40	/public/js/weibo.js?20110824	200	16

表格优化

- 前端美化

- bootstrap

- datable

每页显示条数: 10 ⟷   搜索:

IP	URL	http状态	次数
222.86.153.12	/images/cursor_minify.cur	404	60
222.86.153.12	/images/cursor_zoom.cur	404	32
58.253.6.133	/images/cursor_minify.cur	404	32
111.85.34.165	/%3Ca%20href=	404	28
218.29.111.117	/images/cursor_zoom.cur	404	27
58.253.6.133	/images/cursor_zoom.cur	404	27
218.29.111.117	/images/cursor_minify.cur	404	26
117.63.146.40	/public/js/common.js?20110824	200	19
117.63.146.40	/favicon.ico	404	18
117.63.146.40	/public/js/weibo.js?20110824	200	16

第 1 到 10 条记录，共 107 条

上页 1 2 3 4 5 … 11 下页

简单的可视化

```python
@app.route('/chart')
def chart():
    return render_template('chart.html')
@app.route('/chartdata')
def chartdata():
    sql = 'select status,sum(value) from log group by status '
    res = {
        'label':[],
        'data':[]
    }
    cur = db.execute(sql)
    for c in cur.fetchall():
        res['label'].append(str(c[0]))
        res['data'].append({'name':str(c[0]),'value':int(c[1])})
    return json.dumps(res)
```

```html
<script>
    // 基于准备好的dom, 初始化echarts图表
    var myChart = echarts.init(document.getElementById('main'));
    $.getJSON('/chartdata', function(res) {

        var option = {
            title: {…
            },
            tooltip: {…
            },
            legend: {…
            },
            toolbox: {…
            },
            calculable: true,
            series: [{
                name: '访问数据',
                type: 'pie',
                radius: '55%',
                center: ['50%', '60%'],
                data: res.data
            }]
        };


        myChart.setOption(option);
    })
</script>
```
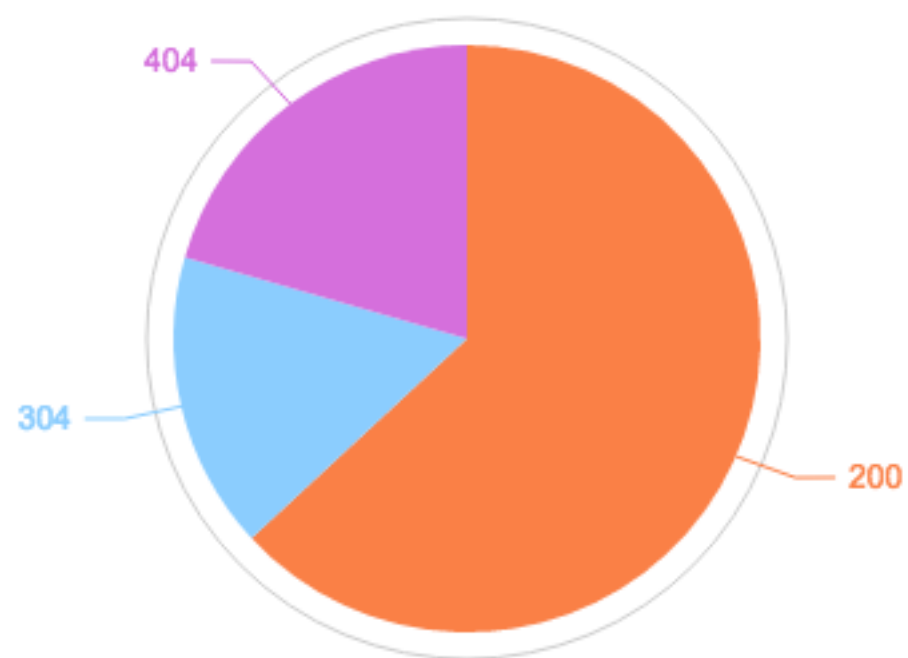
后续分享

- 地理位置

- 访问图

- 需要录屏和ppt的请联系课程顾问

- 课程链接

- QQ ada女神:279312229

- QQ 小夏女神:979950755

谢谢