### 1、采集服务框架

采集框架：



### 2、采集环境目录结构



### 3、采集服务器的启动和关闭

将collect\_srv目录放在/home目录中，作为采集服务根目录（目录名称可以改，但要在编译之前修改）。

##### 1）编译方法

**方法一：**进入采集服务根目录，执行compile\_all.sh，如有出错情况（例如缺少OpenSSL、MySQL的头文件、链接库），根据出错信息安装依赖的文件（使用yum源）。

**方法二：**也可以分开编译，分别执行（./）：

compile\_openresty.sh（编译安装openresty）

compile\_redis.sh（编译安装redis）

compile\_pylib.sh（编译安装第三方Python库，包括一个操作Redis和操作MySQL的库）

compile\_create\_csv.sh（创建导出csv功能）

compile\_keep\_user.sh（留存用户）

在根目录下由如下子目录：

bin:存放编译后的可执行文件

conf:存放配置文件和证书文件，其中Nginx配置文件：nginx\_openresty.conf、Redis配置文件redis.conf和优化配置文件sslproxy.conf文件在后面做性能优化时做相应修改

lib:下载安装过的的源文件

lua\_file:需要执行的lua脚本

sql:无用，可以删除

src:编译目录

##### 2）可能遇到的问题

在编译安装过程中，会遇到需要安装相关第三方开发库的错误，例如，本机安装时报：sh:mysql\_config: command not found

此时，需要安装MySQL对应的开发库，通过下面命令列出MySQL库：

yum list|grep mysql

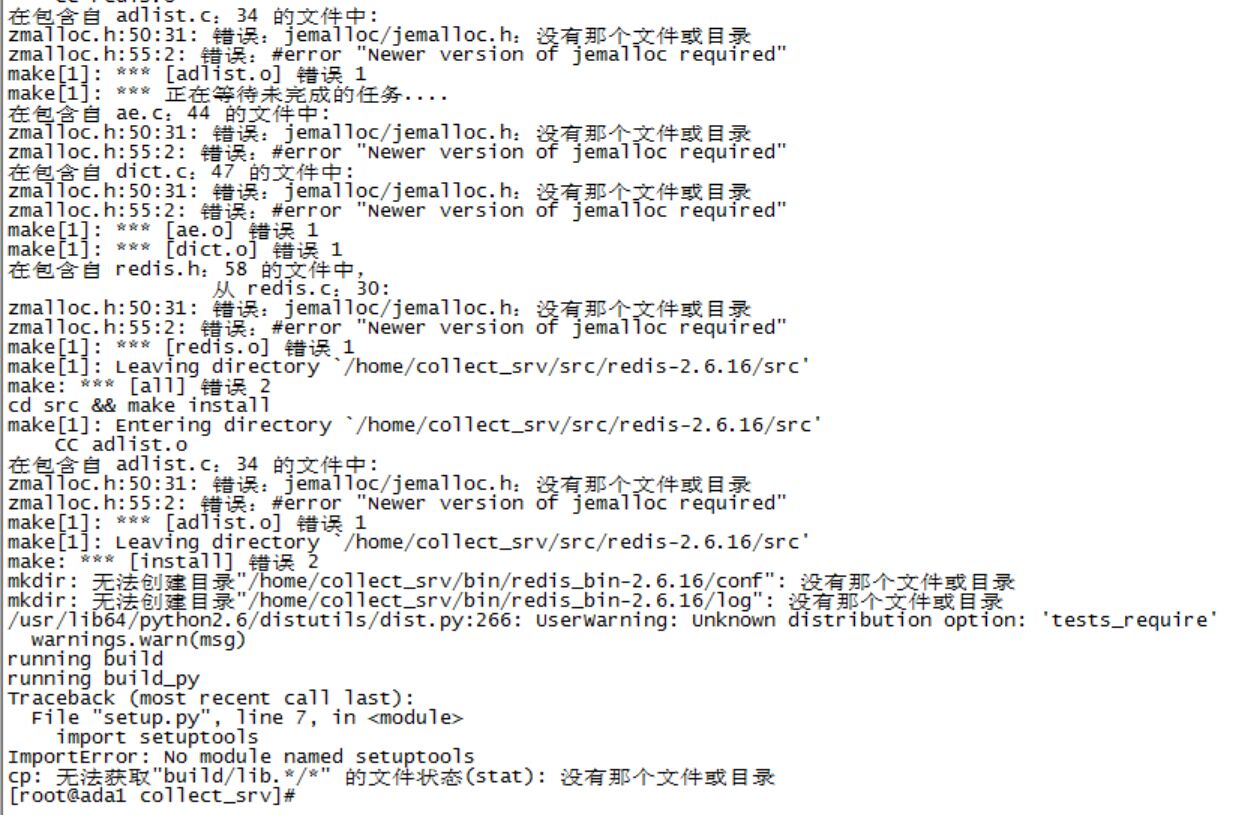
找到对应的开发库，安装：

yum install mysql-devel.x86\_64

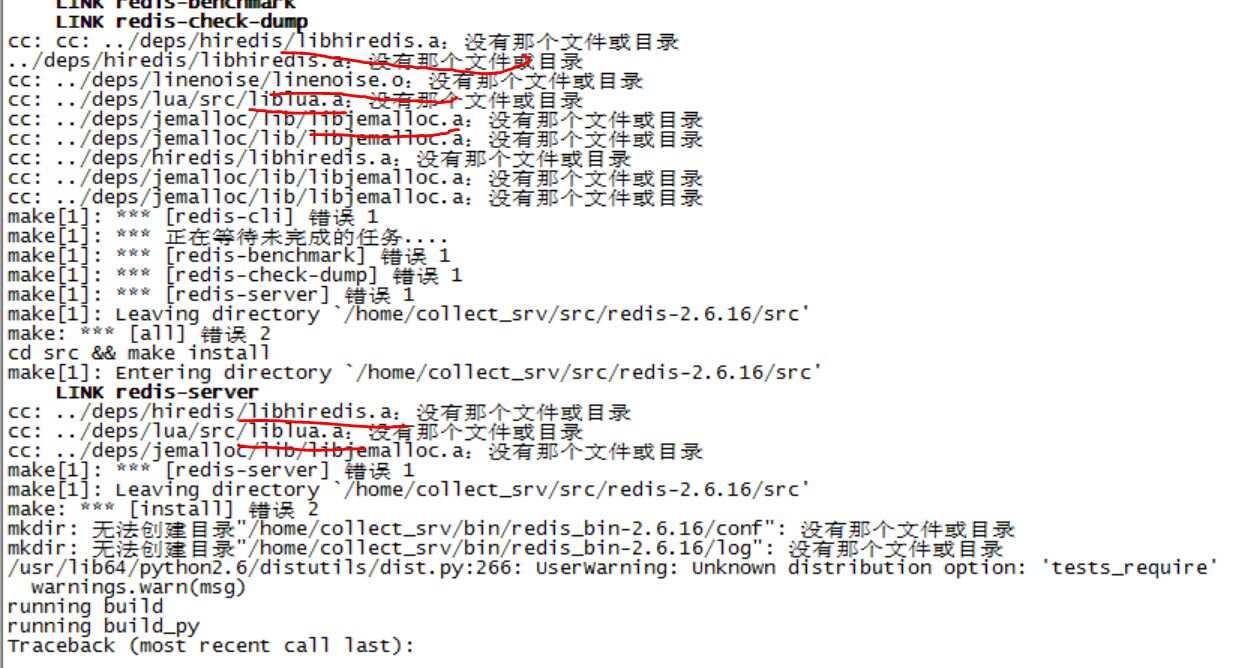
在本次安装过程中还遇到了缺少Python开发库的问题，解决发放同上：

yum install python-devel.x86\_64

wget http://www.canonware.com/download/jemalloc/jemalloc-2.2.5.tar.bz2 .  
tar -jxvf jemalloc-2.2.5.tar.bz2  
cd jemalloc-2.2.5/ && ./configure && make && make install



下载安装jemalloc

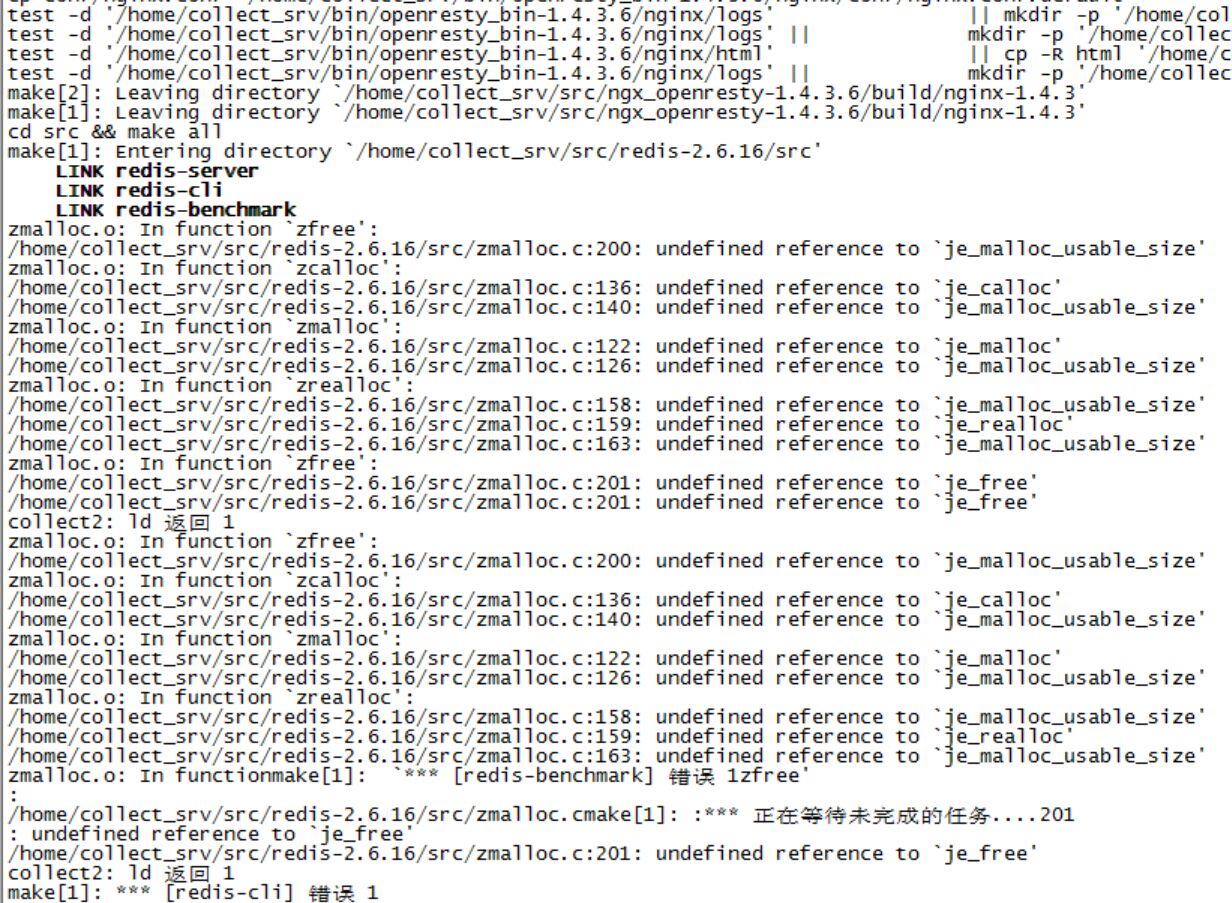


进入src/redis-2.6.16/deps文件夹，

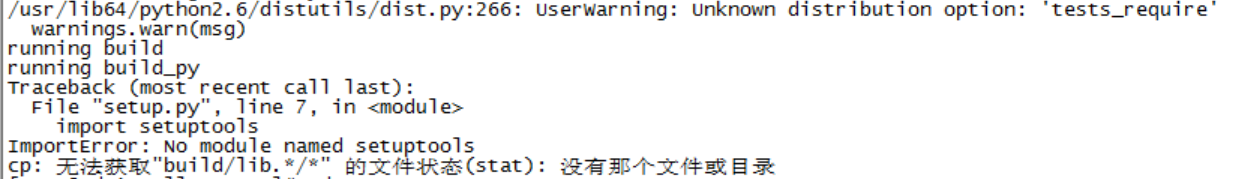
make lua hiredis linenoise

cd ..

make



该问题原因是上面安装的jemalloc版本太老，重新下载安装了jemalloc-3.6.0，删除src目录下的redis-2.6.16，重新编译，redis安装通过。



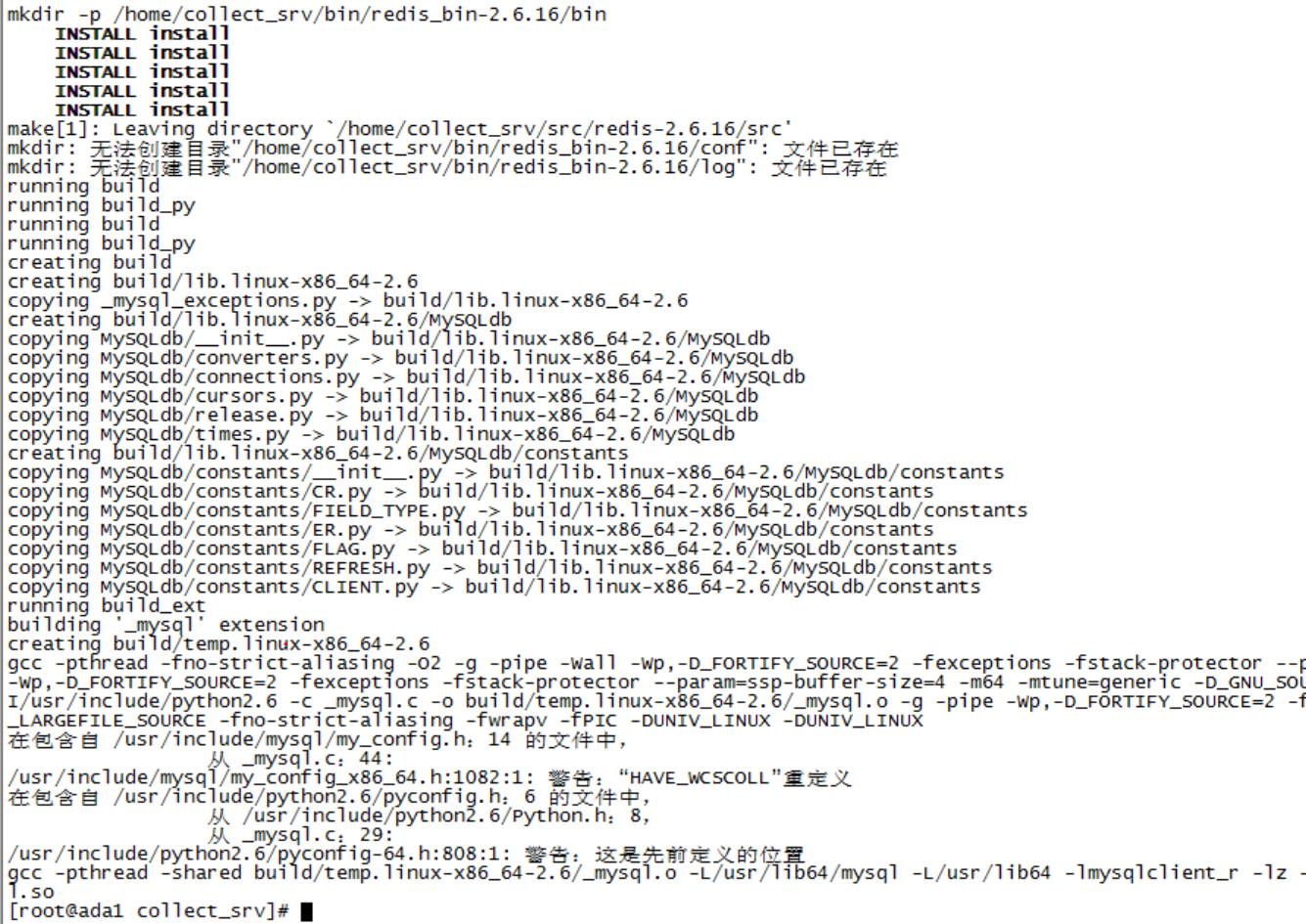
对该问题，安装Python的setuptools，下载setuptools-2.2，进入该文件目录下，执行：

./setup.py build

./setup.py install

再次编译，完成！

上述所需工具包在lib/ 目录下都可以找到！



测试：

服务器ip:192.168.1.211,

Nginx端口：8081

通过chrome的Postman发送数据：

<http://192.168.1.211:8081/app_behavior_analyse>

此处遇到的问题是上面的请求不能得到相应，原因是Linux防火墙设置问题

在Linux防火墙上打开8081端口的input和output：

service iptables -I INPUT -p tcp --dport 8011 -j ACCEPT

service iptables -I OUTPUT -p tcp --sport 8011 -j ACCEPT

service iptables -I INPUT -p udp --dport 8011 -j ACCEPT

service iptables -I OUTPUT -p udp --sport 8011 -j ACCEPT

保存配置：

service iptables save

重启服务：

service iptables restart

### 4、采集服务器的启动和关闭

执行./control.sh start启动采集服务，正常启动后，可以使用ps aux|grep nginx 和 ps aux|grep redis 查看Nginx和Redis进程启用情况。

执行./control.sh stop关闭采集服务。

### 5、修改配置

基本配置可以参考现有配置文件，现在发现的需要修改和注意的配置项如下：

##### 1）nginx\_openresty.conf文件

**work\_process:**该选项控制nginx worker进程数，当前值为1

**worker\_cpu\_affinity：**控制nginx worker进程绑定到对应的cpu核心，当前该选项没用，比如后期根据情况设置为 1000 0100 0010 0001;，表示4cpu情况。

**error\_log：**控制log级别，当前级别为debug，上线后应设置为error级别。

**access\_log:** 在server段中添加access\_log off;关闭

此外，当前Nginx配置文件中，server段中location /add\_dc\_code需保留，/test和/debug可删除。

##### 2）redis.conf文件

**save：**数据持久化配置，后面根据要求配置

##### 3）sslproxy.conf文件

优化配置，比如：

**proxy\_connect\_timeout:** 当前配置为6秒，后面可以适当配长些；

**buff：**缓存配置也需要在优化时根据情况做相应变动

##### 4）连接MySQL配置

需要在/lib/create\_csv目录下的create\_csv.py 和/lib/keep\_user目录下的set\_keepuser\_data.py文件中更改连接的数据库地址并注意使用相应用户。

##### 5）软件版本升级

env.sh中存放软件当前版本号，需要升级时修改版本号，下载新的软件包到lib目录下，重新编译软件。

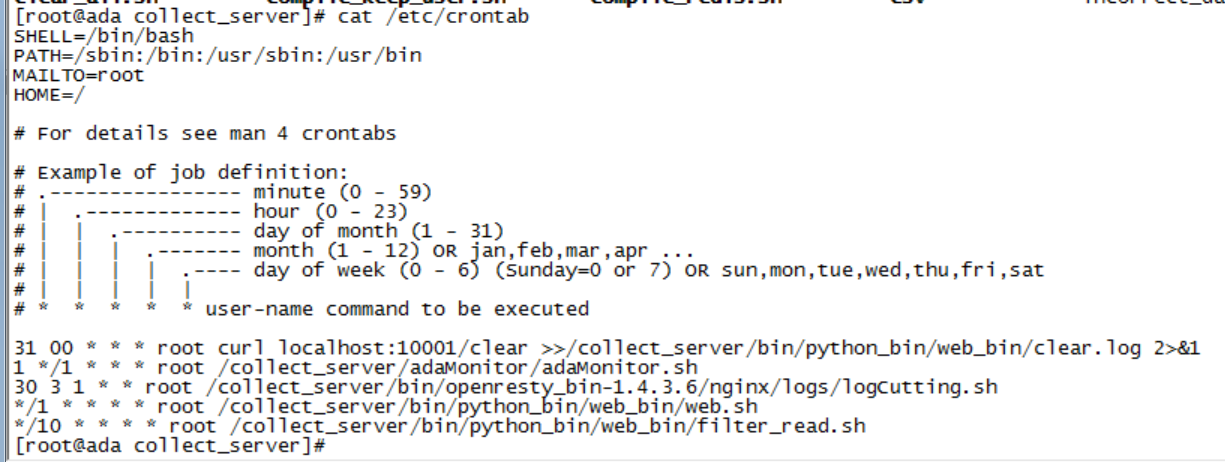
### 6、.csv文件导出配置

当前采集服务和整合层在同一台服务器部署，导出文件存储在根目录下的csv/路径下.

### 7、程序调度

当前程序调度一般使用Linux的定时任务，web采集和app采集当前的定时任务设置如下，具体功能见脚本。

web采集：



app采集：

