邮件查询数据库功能说明

[1. 启动(start.py) 3](#_Toc531789436)

[1.1功能说明 3](#_Toc531789437)

[1.2请求参数 3](#_Toc531789438)

[1.3同步流程 3](#_Toc531789439)

[1.4处理流程 4](#_Toc531789440)

[2. 邮件数据分析(mailsa.py) 4](#_Toc531789441)

[2.1功能说明 4](#_Toc531789442)

[2.2请求参数 5](#_Toc531789443)

[2.3同步流程 5](#_Toc531789444)

[2.4处理流程 5](#_Toc531789445)

[3. 附件分析处理(s2file.py) 6](#_Toc531789446)

[3.1功能说明 6](#_Toc531789447)

[3.2请求参数 6](#_Toc531789448)

[3.3同步流程 6](#_Toc531789449)

[3.4处理流程 7](#_Toc531789450)

[4. 导出数据处理(filepg.py) 8](#_Toc531789451)

[4.1功能说明 8](#_Toc531789452)

[4.2请求参数 8](#_Toc531789453)

[4.3同步流程 8](#_Toc531789454)

[4.4处理流程 9](#_Toc531789455)

[5. 文件分析发送(mailss.py) 10](#_Toc531789456)

[5.1功能说明 10](#_Toc531789457)

[5.2请求参数 10](#_Toc531789458)

[5.3同步流程 10](#_Toc531789459)

[5.4处理流程 11](#_Toc531789460)

[6. 配置文件 12](#_Toc531789461)

[6.1 start.ini 12](#_Toc531789462)

[6.2 paras.csv 13](#_Toc531789463)

# 启动(start.py)

## 1.1功能说明

系统入口，创建多个线程并发处理各自任务。

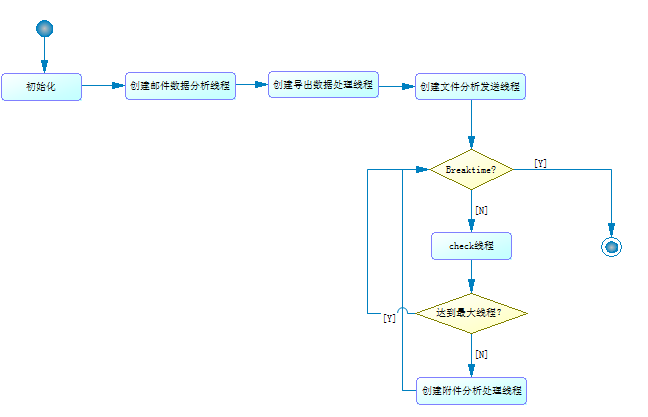
## 1.2请求参数

-v

说明：设定日志数据级别为DEBUG

## 1.3同步流程

流程图说明：



## 1.4处理流程

1.初始化部分：读取系统ini配置文件，设定日志级别，初始化各种目录（下载，日志，历史，发送），刷新db&os配置参数

2.创建邮件数据分析线程：此线程负责读取邮件，下载附件，获取系统运行需要的mailmeta元数据

3.创建导出数据处理线程：此线程负责数据库服务器上已经导出数据文件的压缩&分割&下载&删除

4.创建文件分析发送线程：此线程负责已下载的导出数据的邮件发送，清理历史数据

5.最后一块逻辑是循环创建从数据库导出数据的线程，由于此线程是**瓶颈**，需要创建多个，每个线程处理完自己的事务就退出，如果设定时间没有退出（30Min）,将被系统杀死，然后按照设定参数循环创建新线程。

# 邮件数据分析(mailsa.py)

## 2.1功能说明

按照start.ini的配置参数登录邮件服务器，读取设定目录邮件，获取关键数据，下载附件

## 2.2请求参数

pmailslst,pparaslst,pmutex

说明：

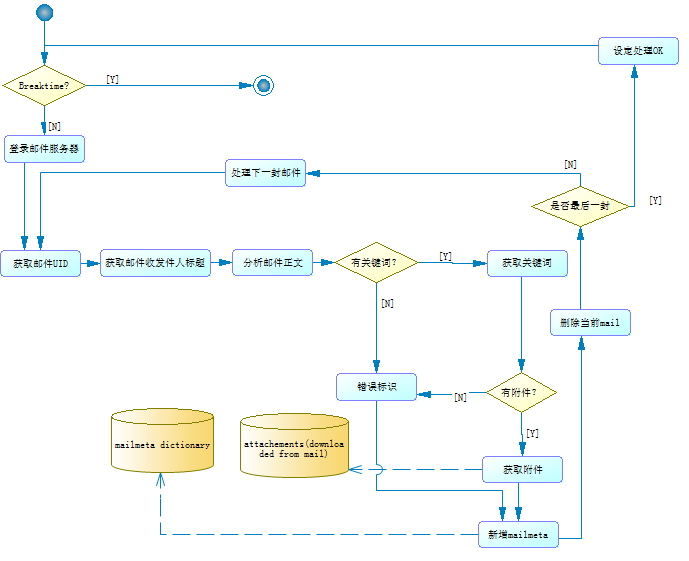
pmailslst:邮件列表,存储从邮件中获取的元数据（收件人，发件人,抄送人，主题，关键词对应的tns），各线程共享

pparaslst:数据库用户名&密码&端口Servername(用于连接数据库执行数据导出)，OS IP&用户名&密码&端口（连接OS执行压缩下载等任务），TNS相关信息（用于sqlplus批处理）；密码加密存储

pmutex:线程锁,用于保护pmailslst，防止出现多线程操作时的数据安全问题

## 2.3同步流程

流程图说明：



## 2.4处理流程

1.判断是否系统休息时间，是则退出

2.按照start.ini配置参数登录邮件服务器

3.获取邮件的唯一性标识UID,此UID也是邮件列表中元数据的唯一标识

4.获取邮件头信息，主要包括发件人，收件人，抄送人，主题

5.分析邮件正文，如果没有关键词database:tnsdescription,标识邮件列表中的元数据为错误

6.判断有无附件，如果没有附件，标识邮件列表中的元数据为错误，有附件则开始下载到配置参数指定的目录

7.添加元数据到邮件列表(maillst)

8.删除当前邮件

9.循环处理下一封邮件，即从步骤3开始

10.如果所有邮件处理结束，设定邮件处理标示为OK,此标识将触发附件分析处理线程开始运行

11.开始下一次循环，即从步骤1开始

# 附件分析处理(s2file.py)

## 3.1功能说明

按照start.ini配置参数从设定目录读取从邮件下载的文件，处理文件成固定格式，调用sqlplus执行批处理创建view,然后通过cx\_oracle连接数据库，调用数据导出存储过程

## 3.2请求参数

pmailslst,pparaslst,pmutex,pdb,pproclst,pmutexp,pdbconlst

说明：

pmailslst:参见邮件数据分析说明

pparaslst:参见邮件数据分析说明

pmutex:参见邮件数据分析说明

pdb:tns标识，此标识限定了当前线程能处理的文件前缀，能登录的数据库及OS

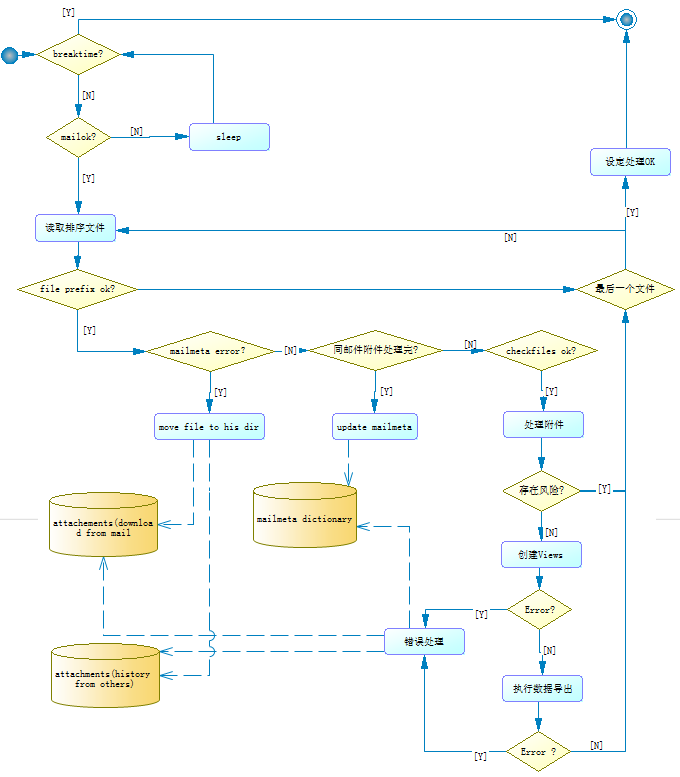
pproclst:进程列表，主要用于存储sqlplus信息，已废弃

pmutexp:线程锁，用于保护pproclst& pdbconlst数据在多线程下的安全

pdbconlst:数据库连接列表，主要用于存储各个线程cx\_oracle模块发起的数据库连接信息

## 3.3同步流程

流程图说明：



## 3.4处理流程

1. 判断是否系统休息时间，是则退出
2. 判断邮件处理是否OK
3. 按照设定目录读取从邮件下载的附件，
4. 判断文件前缀tns信息，是否归本线程处理
5. 判断此文件的UID对应的邮件列表是否已标识错误，如果错误，把文件移动到历史目录，不再处理
6. 判断此文件UID是否和上一个相同，不相同则表示上一邮件的所有附件都处理完，更新上一个UID邮件列表标识为数据已导出
7. 检查文件，判断是否有文件为废弃（可能线程异常或退出导致未处理），有则移动到历史目录，不再继续处理此文件
8. 处理文件，安装sqlplus创建view格式要求处理文件
9. 判断文件是否有风险因子（drop table,truncate table,alter table等），有则标识邮件列表为错误，不再继续处理
10. 调用sqlplus执行批处理，创建view
11. 判断sqlplus是否有错误，有则标识邮件列表为错误，不再继续处理
12. 调用cx\_oracle连接数据库，执行数据导出
13. 判断导出过程是否有错误，有则标识邮件列表为错误，不再继续处理
14. 判断是否最后一个文件，否则循环处理，跳到3处
15. 设定附件分析处理OK,次标识将触发导出数据处理线程
16. 继续下一次循环，跳到1处

# 导出数据处理(filepg.py)

## 4.1功能说明

安照邮件列表中附件处理OK的元数据和配置参数，登录设定服务器，执行压缩，分割，下载导出文件

## 4.2请求参数

pmailslst,pparaslst,pmutex

说明：

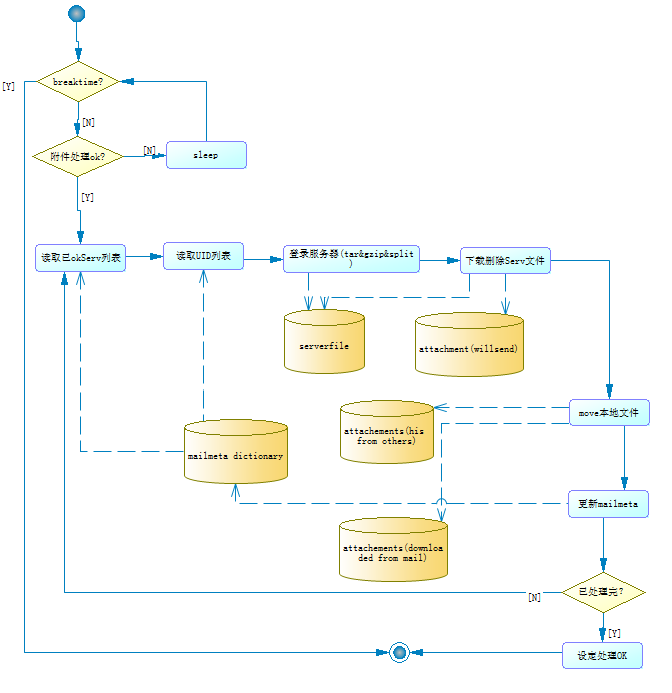
pmailslst:参见邮件数据分析说明

pparaslst:参见邮件数据分析说明

pmutex:参见邮件数据分析说明

## 4.3同步流程

流程图说明：



## 4.4处理流程

1.判断是否系统休息时间，是则退出

2.判断附件处理是否OK

3.读取邮件列表中附件处理OK的服务器(TNS)列表，获取IP&用户&密码&端口

4.循环读取邮件列表中附件处理OK的uid列表，10个一组，按uid列表生成文件合并&压缩&分割shell命令

5.按照第3步读取的参数，登录服务器，上传&执行shell命令文件

6.下载服务器上处理完毕的文件到本地设定文件夹，下载完成后，删除服务器文件

7.移除本地文件到历史目录（已废弃）

8.更新邮件列表中对应uid元数据标识为已下载

9.邮件列表中附件处理OK的服务器(TNS)列表是否已经处理完，否，则循环，跳到第3步

10.设置导出数据处理为OK,此标识将触发文件分析发送线程

11.继续下一次循环，跳到1处

# 文件分析发送(mailss.py)

## 5.1功能说明

读取已下载待发送附件文件夹，根据邮件列表中标识为已下载的元数据，结合配置参数，登录邮件服务器，上载附件文件，发送邮件。

## 5.2请求参数

pmaillst,pmutex,pdirattachment

说明：

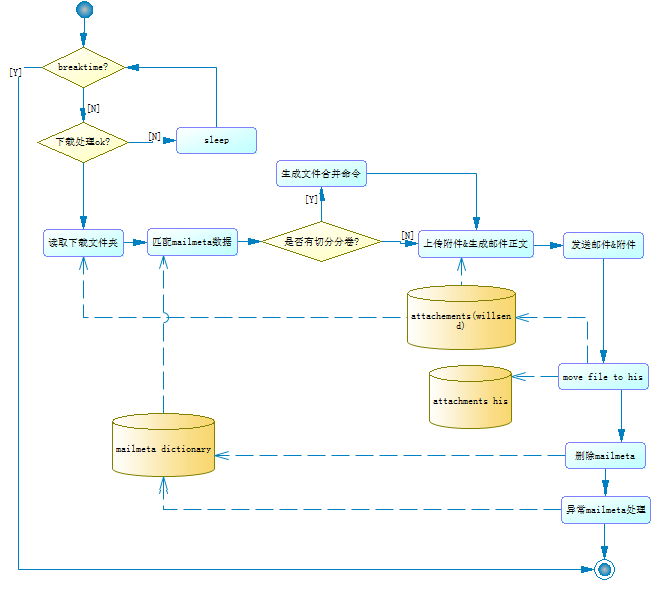
pmaillst:参见邮件数据分析说明

pmutex:参见邮件数据分析说明

pdirattachment:要发送的附件所在目录

## 5.3同步流程

流程图说明：



## 5.4处理流程

1.判断是否系统休息时间，是则退出

2.判断导出数据处理是否OK

3.循环读取已下载待发送的文件夹

4判断是否有分割文件，有则生成合并命令脚本文件

5.上传待发送的文件（如果有多个分割文件，最后一个分割文件将和合并脚本文件一起上传）

6.移除文件到历史目录

7.删除邮件列表中已经发送的uid元数据，如果还有文件待处理，循环，跳到3

8.处理异常邮件列表（主要是超时邮件，包括执行数据导出但超过设定时间未结束，或已经标识为已下载但长时间（最大设定时间\*4）未发送未删除），这些邮件将被标识未错误，发送错误信息后删除。

# 配置文件

## start.ini

系统运行相关配置，实时刷新，更改立即生效，说明如下：

**系统运行参数**

**[run]**

stoptime = 3,4,5,6 #系统休息时间(HH24格式)，逗号隔开，在此设定时间系统自动停止运行且无法开启,如果需要正常停止系统，请增加当前时间到此参数，系统处理完当前事务后自动终止

port=64443 #系统开启时占用端口，主要用于防止系统重复运行

maxthread=3 #瓶颈线程s2file.py最大线程数量

sleep=60 #每一个周期结束后，系统休息时间，此参数决定了系统运行速度，按IPS目前的邮件量此设置不应该低于60，如果系统繁忙，可适当设小

**邮件相关参数**

**[mail]**

user = [Yourname@email.com](mailto:Yourname@email.com) #登录邮件系统的用户名

pwd = #登录邮件系统的明文密码，变更密码时填写此值，系统会自动加密存储于enpwd,并清空此值

enpwd = VGxxhyU= #登录邮件系统的密文密码，系统自动填写

cycle = 600 #重新登录邮件系统的周期（秒）

fromallow = [xianglei.zhu@hotmail.com,name2@emai.com](mailto:xianglei.zhu@hotmail.com,name2@emai.com) #发件人许可

maxqtime = 1800 #每封邮件允许的最大查询运行时间，超时s2file线程将被终止，系统回复超时邮件

**imap邮件协议参数**，mailsa模块使用

**[imap]**

workdir = IPSAutoQuery #当前系统的工作目录

port = 993 #协议端口号

host = imap.xxxx.com #协议host

**smtp邮件协议参数**，mailss模块使用

**[smtp]**

port = 994 #协议端口号

host = smtp.xxxx.com #协议host

**系统目录参数**

**[dir]**

downloaded = D:\AutoQuery\downloaded #邮件附件下载存放目录，mailsa,s2file使用

willsend = D:\AutoQuery\willsend #邮件发送附件目录，filepg,mailss使用

uploaded = D:\AutoQuery\uploaded #需要上传的文件目录，已废弃

history = D:\AutoQuery\his #历史目录，来源于downloaded,willsend

log = D:\AutoQuery\log #系统日志目录，每天一个日志

## paras.csv

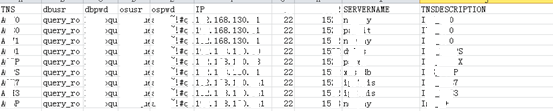
database&os相关配置参数，主要是s2file,filepg模块使用，启动时刷新，为数据安全性考虑，刷新后此文件将被删除，后续使用paras.bin二进制加密文件替代，如果要更新参数，请按照下文格式与顺序修改此文件，放置于源代码目录即可，栏位说明如下：

TNS：系统运行主机上TNSNAME.ORA文件的配置名，注意当前系统使用Oracle 11g client

dbusr&dbpwd&1522&servername(端口):登录数据库用户名&密码&端口&服务名；s2file模块的sqlplus&cx\_oracle使用

ospusr&ospwd&22(端口):ssh&sftp登录os使用的用户名&密码&端口,filepg模块使用

TNSDESCRIPTION:TNS描述，主要用于邮件正文database:关键字，用于区分数据库



created by [xlzhu@ips.com](mailto:xlzhu@ips.com) at Shanghai 2018.12.04