第一步：上传

程式：MAUFILEA03 call MARFILEA01

程式运行方式-Windows自启动

程式logs：ALLIE_UPLOADFILE_LOG

上传前路径（Client端172.30开头IP）：D:\MapData

上传后路径（Server端10.41.158.72）：/data/sftp/aoisftp/upload/

```
--MAUFILEA03 call MARFILEA01 to upload file to SFTP log
--sfcfa139@p8prd
select b.line, b.stage, a.*
--select count(distinct usn)
 from allie_uploadfile_log a, sfcworkstation b
 where a.clientip = b.workstationip
 and a.trndate < sysdate - 1
 and a.trndate > sysdate - 3
 and a.filetype = 'jpg'
  and b.stage in ('FU')
  and a.filename like 'D:\MapData%'
  and a.status = '1'
--sfcfa139@p8prd
--alert mail:主题: UPLOAD AOI CSV&JPG REPORT -- (2020/08/11 09:00:26)
select t.* from sfcalertbase t where t.alertid='191';
select t.* from sfcalertbasesql t where t.alertid='191';
select t.* from sfcalertlogdetail t where t.alertid='191'and t.trndate>sysdate-1;
```

第二步：分拣 by 产品类型 例如Carrier和Keyboard

1.将正常JPG文件分拣至目录/data/buffer/total_jpg/

2.将carrier JPG分拣至目录/data/buffer/total_jpg_carrier/

3.将异常文件分拣至目录/data/buffer/err_jpg/

程式（Server端10.41.158.72）：/home/hadoop/program/total_scripts/01jpg_migrate_20190423_v1.py

程式运行方式-Linux的crontab

程式logs：无

分拣前路径：/data/sftp/aoisftp/upload/

分拣后OK路径：/data/buffer/total_jpg/

分拣后carrier_jpg路径：/data/buffer/total_jpg_carrier/

分拣后NG路径：/data/buffer/err_jpg/（有用Crontab定期删除文件）

查找定期任务命令：crontab -l | grep -n err_jpg

49:5 7 * * * /usr/bin/find /data/buffer/err_jpg/ -mtime +7 -name "*" -exec rm -f {} \;

每隔7天的5点49分执行Crontab，

执行内容是删除/usr/bin/find /data/buffer/err_jpg/ 目录下面所有超过7天的文件

```
[hadoop@zsp8kamap02 /]$ pwd
/
[hadoop@zsp8kamap02 /]$ crontab -l | grep -n err_jpg
```

```
49:5 7 * * * /usr/bin/find /data/buffer/err_jpg/ -mtime +7 -name "*" -exec rm -f {} \;
[hadoop@zsp8kamap02 /]$
```

```
[mike@zsp8kamap02 upload]$ cd /data/sftp/aoisftp
[mike@zsp8kamap02 aoisftp]$ pwd
/data/sftp/aoisftp
[mike@zsp8kamap02 aoisftp]$ ls -lR | grep "FQC-LEGEND"| wc -l
2281
[hadoop@zsp8kamap02 upload]$ ls -lR | find -name '*FQC-LEGEND*'| wc -l
1898
[mike@zsp8kamap02 aoisftp]$ cd /data/history/
[mike@zsp8kamap02 history]$ pwd
 /data/history
 [mike@zsp8kamap02 history]$ ls -lR | grep "FQC-LEGEND"| wc -l
 0
 [mike@zsp8kamap02 history]$ cd /data/buffer/err_jpg/
 [mike@zsp8kamap02 err_jpg]$ pwd
 /data/buffer/err_jpg
 [mike@zsp8kamap02 err_jpg]$ ls -lR | grep "FQC-LEGEND"| wc -l
 15190
 [hadoop@zsp8kamap02 err_jpg]$ ls -lR | find -name '*FQC-LEGEND*'| wc -l
 15190
 [mike@zsp8kamap02 err_jpg]$
```

第三步：分拣 by 日期 例如 今日 和 非今日

1.将非今日文件分拣至目录/hdata/buffer/debug_total_jpg/today/

2.将今日文件分拣至目录/data/buffer/total_jpg/

程式（Server端

10.41.158.72）：/home/hadoop/program/total_scripts/00jpg_migrate_today_before_20190111_v1.py

程式运行方式-Linux的crontab

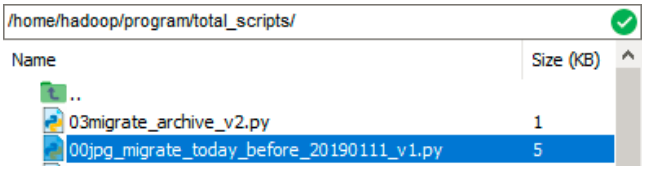
程式logs：无

分拣前路径：/data/buffer/total_jpg/

将非今日文件分拣至目录/hdata/buffer/debug_total_jpg/today/

将今日文件分拣至目录/data/buffer/total_jpg/

分拣后carrier_jpg路径：/data/buffer/total_jpg_carrier/



第四步：分拣 by Loading 例如 将非今日的文件均摊分存到6个文件夹

将非今日目录内文件随机放入6个文件夹

/data/buffer/debug_total_jpg/100

/data/buffer/debug_total_jpg/600

程式（Server端
10.41.158.72）：/home/hadoop/program/01jpg_errorfile_modify_debug_total_jpg_today_v1.py
程式运行方式-Linux的crontab
程式logs：无
分拣前路径：/data/sftp/aoisftp/upload/
将非今日目录内文件随机放入6个文件夹
/data/buffer/debug_total_jpg/100
…
/data/buffer/debug_total_jpg/600

第五步：上传到HDFS
将JPG文件上传至HDFS，目录/P8AOI/MapData/年月日目录,同时URL信息出入Kudu table:allie.aoi_imageurl Hbase:ie4_p8_csv 一共6只，每只对应一个目录，处理完后存入目录/hdata/buffer/total_pre_arch/
程式（Server端
10.41.158.106）：/home/hadoop/program/total_scripts/02jpg_upload_hdfs_20180607_debug_single_v1_01.py
程式运行方式-gocron
gocron JOB名:jpg_upload_single01~06

非今日
将今日文件按照目录上传至HDFS，目录/P8AOI/MapData/年月日目录,同时URL信息出入Kudu table:allie.aoi_imageurl Hbase:ie4_p8_csv 一共3*6=18只，每只程序每6小时循环一次，每只对应一个源目录，临时目录
程式（Server端
10.41.158.63）：/home/hadoop/program/total_scripts/02jpg_upload_hdfs_20180608_v1_debug_1.py

临时目录
/bfdata/buffer/total_tmp01~06
/bfdata/buffer/total_tmp11~16
/bfdata/buffer/total_tmp21~26
处理完后存入目录
/hdata/buffer/total_pre_arch/
JOB执行方式-crontab

第六步：归档
将待封存文件夹下文件按照日期封存于下面目录，待效验程序使用
/data/history/jpeg/年月日
程式（Server端10.41.158.72）：/home/hadoop/program/total_scripts/03migrate_archive_v2.py
程式运行方式-Linux的crontab
程式logs：无
归档前路径：/hdata/buffer/total_pre_arch/
归档后路径：/data/history/jpeg/年月日

第七步：复检

检查/data/history/jpeg/年月日 中文件是否存在于hdfs目录中，如不存在将文件放回
/hdata/history/jpeg/lost/ 目录

程式（Server端10.41.158.106）：/home/hadoop/program/total_scripts/different.py

程式运行方式-gocron

gocron JOB名:check_JPG_compare

程式logs：无

归档前路径：/data/history/jpeg/年月日

归档后路径：/hdata/history/jpeg/lost/ 目录

第八步：重新存储lost的JPG到HDFS

将未正确上传HDFS的JPG文件从文件夹/data/history/jpeg/lost/移动
到/hdata/buffer/debug_total_jpg/100~600

程式（Server端
10.41.158.72）：/home/hadoop/program/01jpg_errorfile_modify_debug_total_jpg_lost_v1.py

程式运行方式-Linux的crontab

程式logs：无

归档前路径：/hdata/history/jpeg/lost/ 目录

归档后路径：/hdata/buffer/debug_total_jpg/100~600

第九步：定期清理 by 时间（计划后续增加机种条件做判断）

从HDFS删除超一年的JPG文件，同步清理Hbase 资料

程式（Server端10.41.158.106）：/home/hadoop/program/DelJPG_20170811_v1.py

程式运行方式-gocron

gocron JOB名:Del_1Y_JPG

程式logs：无

归档前路径：/hdata/history/jpeg/lost/ 目录

归档后路径：/hdata/buffer/debug_total_jpg/100~600

第十步：查询和下载

HUE：http://p8cdhmasterp01.wzs.wistron:8888/hue/editor?editor=45860

```
SELECT * from allie.aoi_imageurl
where imageurl like'%FQC%'--查询FU站的
and trndate>seconds_sub(now(), 10)--10秒前
```

HDFS

```
[hadoop@p8aoiprd06 ~]$ hdfs dfs -du /P8AOI/MapData/com15971/X1777-ANSI-FQC-LEGEND_com15971_TB1-F10-TRI-11@com15971828452020081205514l-FPW033326GBN7Y8A1.JPG
20/08/13 09:20:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1587339 /P8AOI/MapData/com15971/X1777-ANSI-FQC-LEGEND_com15971_TB1-F10-TRI-11@com15971828452020081205514l-FPW033326GBN7Y8A1.JPG
[hadoop@p8aoiprd06 ~]$ hdfs dfs -du -h /P8AOI/MapData/com15971/X1777-ANSI-FQC-LEGEND_com15971_TB1-F10-TRI-11@com15971828452020081205514l-FPW033326GBN7Y8A1.JPG
20/08/13 09:21:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
1.5 M /P8AOI/MapData/com15971/X1777-ANSI-FQC-LEGEND_com15971_TB1-F10-TRI-11@com15
718284520200812055141-FPW033326GBN7Y8A1.JPG
[hadoop@p8aoiprd06 ~]$
```

```
删除文件夹
hdfs dfs -rm -r /P8AOI/MapData/X530-PSA-JIS
删除文件
hdfs dfs -rm /P8AOI/MapData/X530-PSA-JIS
列清单
hdfs dfs -ls /P8AOI/MapData/20200813/
[hadoop@p8aoiprd06 ~]$ hdfs dfs -ls /P8AOI/MapData/20190813/
20/08/13 09:26:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
ls: `/P8AOI/MapData/20190813/': No such file or directory
[hadoop@p8aoiprd06 ~]$

```

HBASE-jpg查询

```
[mike@p8aoiprd06 ~]$ su - hadoop
Password:
[hadoop@p8aoiprd06 ~]$ hbase shell
2020-08-13 09:14:07,500 WARN [main] util.NativeCodeLoader: Unable to load native-
adoop library for your platform... using builtin-java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/wistron-hadoop/hbase-1.2.6/lib/slf
j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/wistron-hadoop/hadoop-2.7.1/share/
adoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

hbase(main):001:0> desc 'ie4_p8_aoi'
Table ie4_p8_aoi is ENABLED
ie4_p8_aoi
COLUMN FAMILIES DESCRIPTION
{NAME => 'cf', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_
DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL
 => '31536000 SECONDS (365 DAYS)', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BL
OCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE
=> '0'}
1 row(s) in 0.4630 seconds

hbase(main):001:0> get 'ie4_p8_aoi','FPW033326GBN7Y8A1',{COLUMNS=>'cf'}
```

```
SyntaxError: (hbase):1: syntax error, unexpected tIDENTIFIER

get 'ie4_p8_aoi','FPW033326GBN7Y8A1',{COLUMNS=>''cf}
 ^

hbase(main):002:0> get 'ie4_p8_aoi','FPW033326GBN7Y8A1',{COLUMNS=>'cf'}
COLUMN CELL
 cf:20200811220345 timestamp=1597157381524, value=/P8AOI/MapData/20200811/X1777-A
NSI-MB-180_20200811_TB1-F10-TRI-02@2020081
 1220345-FPW033326GBN7Y8A1.JPG
 cf:20200811222034 timestamp=1597157402160, value=/P8AOI/MapData/20200811/X1777-A
NSI-TOP-180_20200811_TB1-F10-TRI-03@202008
 11222034-FPW033326GBN7Y8A1.JPG
 cf:20200811222605 timestamp=1597157385405, value=/P8AOI/MapData/20200811/X1777-A
NSI-BOT_20200811_TB1-F10-TRI-05@2020081122
 2605-FPW033326GBN7Y8A1.JPG
 cf:20200812035723 timestamp=1597179064837, value=/P8AOI/MapData/20200812/X1777-A
NSI-BOT_20200812_TB1-F10-TRI-05@2020081203
 5723-FPW033326GBN7Y8A1.JPG
 cf:20200812040239 timestamp=1597179020994, value=/P8AOI/MapData/20200812/X1777-A
NSI-HT_20200812_TB1-F10-TRI-07@20200812040
 239-FPW033326GBN7Y8A1.JPG
 cf:20200812055049 timestamp=1597183758105, value=/P8AOI/MapData/20200812/X1777-A
NSI-BOT_20200812_TB1-F10-TRI-09@2020081205
 5049-FPW033326GBN7Y8A1.JPG
 cf:20200812055141 timestamp=1597183743517, value=/P8AOI/MapData/20200812/X1777-A
NSI-FQC-LEGEND_20200812_TB1-F10-TRI-11@202
 00812055141-FPW033326GBN7Y8A1.JPG
 cf:com1597182845202008120551 41 timestamp=1597279713340, value=/P8AOI/MapData/cor
15971/X1777-ANSI-FQC-LEGEND_com15971_TB1-F10-TRI-11@com
 15971828452020081205514 1-FPW033326GBN7Y8A1.JPG
 cf:com1597183183202008120550 49 timestamp=1597278619896, value=/P8AOI/MapData/cor
15971/X1777-ANSI-BOT_com15971_TB1-F10-TRI-09@com1597183
 18320200812055049-FPW033326GBN7Y8A1.JPG
9 row(s) in 0.2190 seconds

hbase(main):003:0>
```

HBASE-csv查询

```
get 'p8_aoi_csv','FPW637460GWHC3F7A',{COLUMNS=>'KN'}

get 'p8_aoi_csv','FPW70761XJ8HGX22G,{COLUMNS=>''WF}
在表'p8_aoi_csv' 中查找序列号'FPW637460GWHC3F7A'
get 'p4_aoi_csv','FPW70761XJ8HGX22G'
scan 'p8_aoi_csv',{COLUMNS=>'KN',LIMIT=>10}
列出表表'p8_aoi_csv' 中columns 为KN 的前10条记录

```

```
9  count 'p8_aoi_csv',{COLUMNS=>'KN'}
10  统计'p8_aoi_csv'中columns为KN的记录数
11  一般能显示行中的序列号
```