# 公安局

1. 前置库

|  |
| --- |
| 数据库类型：mysql  连接服务器：172.26.16.90  连接命令：mysql -h15.75.8.103 -udt -p1Qaz2wsx@  使用dsjzx用户：use dsjzx;  展示所有的表：show tables;  展示某张表的建表信息：show create table base\_crjdata\_cjsqxx\_df; |

1. 生成工作流信息

|  |
| --- |
| {"UnitId":"223372764705329152","UnitName": "gonganju","XinghuanAddress":"jdbc:hive2://172.17.148.189:30925/gonganju\_txt\_ods;guardianToken=YRzeh1lLQQETtBMYg5Sd-851RZ52.TDH","HttpfsAddress": "http://172.17.148.189:32228/webhdfs/v1/tmp/detuo?op=open&guardian\_access\_token=YRzeh1lLQQETtBMYg5Sd-851RZ52.TDH","SourceTableName":"dsjzx.base\_shlg\_t\_hotel\_df","XinghuanTableName":"gonganju\_txt\_ods.base\_shlg\_t\_hotel\_df","MappingList":["h\_lvguandaima","h\_lvguanmingcheng",],"Exectimes":-1,"WorkflowTime":{"ScheduleType":1,"ScheduleDetail":"05:30"},"AddTimefield":"jhpt\_update\_time","PrimaryKey":"empty","XinghuanOrcTableName":"gonganju\_orc\_ods.base\_shlg\_t\_hotel\_df","ColumnType":"date","TargetList":["h\_lvguandaima","h\_lvguanmingcheng"],"Execrate":2}参考：[接口文档](#_接口文档) |

1. 通过接口下发工作流

|  |
| --- |
| 发起post请求：<http://172.26.16.90:12306/dana/workflow/add（新增）>    删除工作流：<http://172.26.16.90:12306/dana/workflow/delete> |

1. 生成DataX数据抽取配置文件(172.26.16.90)

|  |
| --- |
| 全量抽取：/opt/datatom/dana\_api/execute\_all/{tableName}.json  备注：生成配置文件后需更改文件内where条件并手动执行抽取全量数据。  增量抽取：/opt/datatom/dana\_api/execute\_cfg/{tableName}.json  备注：DS平台每日定时抽取增量数据（T-2天的数据）。 |

1. 上传数据至星环hdfs并同步数据至星环hive(172.26.16.90)

|  |
| --- |
| 全量抽取：/opt/datatom/dana\_api/execute\_all/{tableName}.py  备注：python {tableName}.py 会解析json文件进行数据上传。  增量抽取：/opt/datatom/dana\_api/execute\_cfg/{tableName}.py  备注：DS平台每日定时抽取增量数据（T-2天的数据）。 |

# 人社局

1. 前置库数据抽取流程....（后几步参考公安数据抽取）

|  |
| --- |
| 通过datatrip工具连接（如下图）  数据库类型：dameng  前置机服务器：10.81.68.20  端口：5236  用户名：SYSDBA  密码：SYSDBA |

1. ftp文件到本地解析上传至星环

|  |
| --- |
| 1. 登录ftp下载文件   ftp  open 10.81.68.20  用户密码：sbftp/shsb@1234  下载ftp文件到本地  prompt off 关闭下载文件验证  切换到数据目录  mget \* /opt/datatom/xinghuansftp/rensheju/data/   1. 修改配置文件   cd /opt/datatom/xinghuansftp/rensheju/conf  vi info 指定数据目录，星环连接信息     1. 执行python脚本，抽取数据至星环txt   Data目录下文件    执行python文件  cd /opt/datatom/xinghuansftp/rensheju/script  Python rensheju\_month.py   1. 加载数据到星环orc   Insert into orctable select \* from txttable |

# 医保局

1. 前置库

|  |
| --- |
| 通过datatrip工具连接（如下图）    数据库类型：dameng  前置机服务器：10.81.68.15  端口：5236  用户名：SYSDBA  密码：SYSDBA |

1. 通过csv文件上传至星环

|  |
| --- |
| 1. [创建星环csv表](#_Csv表) 2. 将原始csv文件删除第一行及最后一列 并将引号去掉后 将csv文件上传至星环hdfs   curl -i -C - -X PUT -T /opt/datatom/xinghuansftp/yibaoju/7yiliaoqixie.csv "http://172.17.148.189:31734/webhdfs/v1/tmp/detuo/7yiliaoqixie.csv?op=CREATE&data=true&guardian\_access\_token=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH" -H "Content-Type:application/octet-stream"   1. 加载hdfs上的数据到csv表   load DATA inpath '/tmp/detuo/7yiliaoqixie.csv' INTO TABLE `yibaoju\_txt`.`MATBUYINFO\_csv`   1. 将时间字段格式化，结果数据重新写入新的csv表   create table yibaoju\_txt.MATBUYINFO\_csv\_1 as  select id,qybm,qymc,FROM\_UNIXTIME(to\_unix\_timestamp(jhpt\_update\_time,'yyyy/MM/dd HH:mm:ss')) jhpt\_update\_time,jhpt\_deleted,  FROM\_UNIXTIME(to\_unix\_timestamp(addtime,'yyyy/MM/dd HH:mm:ss')) addtime FROM yibaoju\_txt.MATBUYINFO\_csv  ID,HCTBDM,NY,YYBM,QYBM,CGSL,CGJE,JHPT\_UPDATE\_TIME,JHPT\_DELETED,ADDTIME  create table yibaoju\_txt.MATBUYINFO\_csv\_1 as  select ID,  HCTBDM,  NY,  YYBM,  QYBM,  CGSL,  CGJE,  FROM\_UNIXTIME(to\_unix\_timestamp(JHPT\_UPDATE\_TIME, 'yyyy/MM/dd HH:mm:ss')) JHPT\_UPDATE\_TIME,  JHPT\_DELETED,  FROM\_UNIXTIME(to\_unix\_timestamp(ADDTIME, 'yyyy/MM/dd HH:mm:ss')) ADDTIME  FROM yibaoju\_txt.MATBUYINFO\_csv |

# 热线12345

参考医保局CSV数据入湖方式；

# 税务局

目前是通过sftp的方式将数据拉到本地解析入湖

|  |
| --- |
| 星环sftp  sftp [detuo@172.26.41.27](mailto:detuo@172.26.41.27)  1qaz2wsx  Asdf3.14  程序：172-26-16-90:/opt/datatom/myxie/shuiwuju/upload.sh |

[设置多字段分隔符建表](#_设置多字段分隔符建表)

# 司法局

|  |
| --- |
| 司法局相关连接信息 |

# 档案局

|  |
| --- |
| 档案局sftp  172.26.41.27 /home/dft/danganju0903\_csv  sftp detuo@172.26.41.27 1qaz2wsx  数据与表的对应关系  BY---T\_AR\_0100\_BASE  FYTW--T\_AR\_0090\_BASE  SX--T\_AR\_0120\_BASE  XJ--T\_AR\_0080\_BASE  ZSY--T\_AR\_0010\_BASE  GSRD--T\_AR\_0020\_BASE  RCYJ--T\_AR\_0110\_BASE |

处理脚本：/opt/datatom/xinghuansftp/danganju/script/trans.sh

转换文件编码及将csv前2行删除

|  |
| --- |
| **注意：**   1. **“,”**可以有效的避免csv文件中某一列存在逗号的而导致分列 2. Wc -l 命令在统计csv文件数量的时候统计可能存在不准确的情况，因为本身存在换行符，可以用wps打开csv文件（office和wps也可能存在数量对不上的情况） |

|  |
| --- |
| **注意：**   1. **时间字段需要进行格式转换**   *FROM\_UNIXTIME*(*to\_unix\_timestamp*(**jhpt\_update\_time**,**'yyyy-MM-dd HH:mm'**)) jhpt\_update\_time   1. **插入数据时存在换行/回车符，导致插入数据变多**   *replace*(*replace*(**archives\_code**,*chr*(10),**''**),*chr*(13),**''**) |

# 附件

## 接口文档



## 星环相关连接信息

委办局对应hdfs链接



|  |
| --- |
| 星环**公安**jdbcurl：  jdbc:hive2://172.17.148.189:30925/gonganju\_txt\_ods;guardianToken=YRzeh1lLQQETtBMYg5Sd-851RZ52.TDH  星环**公安**httpfs：  [http://172.17.148.189:32228/webhdfs/v1/tmp/?op=open&guardian\_access\_token=YRzeh1lLQQETtBMYg5Sd-851RZ52.TDH](http://172.17.148.189:32228/webhdfs/v1/tmp/detuo?op=open&guardian_access_token=YRzeh1lLQQETtBMYg5Sd-851RZ52.TDH)  星环**税务**jdbcurl：  jdbc:hive2://172.17.148.189:32474/shuiwuju\_txt;guardianToken=eINKm6FvmYc8zBCFi61j-TR8DW0L.TDH  星环**税务**httpfs：  http://172.17.148.189:32370/webhdfs/v1/tmp/detuo?op=open&guardian\_access\_token=eINKm6FvmYc8zBCFi61j-TR8DW0L.TDH  星环**人社**jdbcurl：  jdbc:hive2://172.17.148.189:31326/;guardianToken=GRtS9d0VzDWz4krYZ7ak-M8IUIMT.TDH  星环**人社**httpfs：（每个租户一个地址，除公安和税务外其他委办局公用一个）  http://172.17.148.189:31734/webhdfs/v1/tmp/detuo?op=open&guardian\_access\_token=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH  星环**医保**jdbcurl：  jdbc:hive2://172.17.148.189:31326/yibaoju\_txt;guardianToken=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH  星环**医保**httpfs：（每个租户一个地址，除公安和税务外其他委办局公用一个）  http://172.17.148.189:31734/webhdfs/v1/tmp/detuo?op=open&guardian\_access\_token=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH  星环**热线12345**jdbcurl：  jdbc:hive2://172.17.148.189:31326/rexian\_orc;guardianToken=8OaTwGAflRSKjKEex8gw-M8IUIMT.TDH  星环**热线12345**httpfs：（每个租户一个地址，除公安和税务外其他委办局公用一个）  <http://172.17.148.189:31734/webhdfs/v1/tmp/detuo?op=open&guardian_access_token=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH>  星环**档案**jdbcurl：  jdbc:hive2://172.17.148.189:31326/danganju\_txt;guardianToken=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH  星环**档案**httpfs：（每个租户一个地址，除公安和税务外其他委办局公用一个）  http://172.17.148.189:31734/webhdfs/v1/tmp/detuo?op=open&guardian\_access\_token=IdnsKV9cgCVkfF00VEka-M8IUIMT.TDH  星环**Detuo测试**jdbcurl：  jdbc:hive2://172.26.41.32:30366/tdt;guardianToken=6mKYTUMtterzxlYyMpv4-CEIJAZ5.TDH  星环**Detuo测试**httpfs：  <http://172.26.41.32:32055/webhdfs/v1/tmp/?op=CREATE&guardian_access_token=6mKYTUMtterzxlYyMpv4-CEIJAZ5.TDH>  **司法kundb：**  账户:yiwangtongban  密码:123  IP:172.17.148.185  端口:15307  数据库:yiwangtongban  mysql -h172.17.148.185 -P15307 -uyiwangtongban -p123 --enable-cleartext-plugin |

## 数据库连接相关

### Stork

|  |
| --- |
| source\_info：存储前置机连接信息，ip、端口、库名、用户名、密码等，此表需手动维护;  ob\_run\_info：存储datax日志信息，提供给万达  token\_info :刷新token，并存储在该表内，必须通过接口调用，才会记录  task\_info :记录数据流上下线，必须通过接口调用，才会记录 |

### 万达--委办局id

Mysql数据库

万达资源库 172.26.16.200 33306 longmarch/longmarch

## 星环hdfs操作

|  |
| --- |
| 查看目录：  curl -X GET "http://172.26.41.32:32055/webhdfs/v1/tmp/datatom/?op=LISTSTATUS&guardian\_access\_token=6mKYTUMtterzxlYyMpv4-CEIJAZ5.TDH"  在hdfs上新建目录（文件）：  curl -X PUT "http://172.17.148.189:31734/webhdfs/v1/tmp/datatom/?op=MKDIRS&permission=777&guardian\_access\_token=EAucQR4SuHCcVXt10VeG-M8IUIMT.TDH"  删除目录：  curl -X DETELE "http://172.17.148.189:31734/webhdfs/v1/tmp/datatom?op=DELETE&guardian\_access\_token=EAucQR4SuHCcVXt10VeG-M8IUIMT.TDH"  参考文档： |

|  |
| --- |
| **CURL命令：**  curl是一个命令行工具，通过指定的URL来上传或下载数据，并将数据展示出来。curl中的c表示client，而URL，就是URL。  **curl语法及选项**  在以下选项中，(H) 表示仅适用 HTTP/HTTPS ，(F) 表示仅适用于 FTP  --anyauth 选择 "any" 认证方法 (H)  -a, --append 添加要上传的文件 (F/SFTP)  --basic 使用HTTP基础认证（Basic Authentication）(H)  --cacert FILE CA 证书，用于每次请求认证 (SSL)  --capath DIR CA 证书目录 (SSL)  -E, --cert CERT[:PASSWD] 客户端证书文件及密码 (SSL)  --cert-type TYPE 证书文件类型 (DER/PEM/ENG) (SSL)  --ciphers LIST SSL 秘钥 (SSL)  --compressed 请求压缩 (使用 deflate 或 gzip)  -K, --config FILE 指定配置文件  --connect-timeout SECONDS 连接超时设置  -C, --continue-at OFFSET 断点续转  -b, --cookie STRING/FILE Cookies字符串或读取Cookies的文件位置 (H)  -c, --cookie-jar FILE 操作结束后，要写入 Cookies 的文件位置 (H)  --create-dirs 创建必要的本地目录层次结构  --crlf 在上传时将 LF 转写为 CRLF  --crlfile FILE 从指定的文件获得PEM格式CRL列表  -d, --data DATA HTTP POST 数据 (H)  --data-ascii DATA ASCII 编码 HTTP POST 数据 (H)  --data-binary DATA binary 编码 HTTP POST 数据 (H)  --data-urlencode DATA url 编码 HTTP POST 数据 (H)  --delegation STRING GSS-API 委托权限  --digest 使用数字身份验证 (H)  --disable-eprt 禁止使用 EPRT 或 LPRT (F)  --disable-epsv 禁止使用 EPSV (F)  -D, --dump-header FILE 将头信息写入指定的文件  --egd-file FILE 为随机数据设置EGD socket路径(SSL)  --engine ENGINGE 加密引擎 (SSL). "--engine list" 指定列表  -f, --fail 连接失败时不显示HTTP错误信息 (H)  -F, --form CONTENT 模拟 HTTP 表单数据提交（multipart POST） (H)  --form-string STRING 模拟 HTTP 表单数据提交 (H)  --ftp-account DATA 帐户数据提交 (F)  --ftp-alternative-to-user COMMAND 指定替换 "USER [name]" 的字符串 (F)  --ftp-create-dirs 如果不存在则创建远程目录 (F)  --ftp-method [MULTICWD/NOCWD/SINGLECWD] 控制 CWD (F)  --ftp-pasv 使用 PASV/EPSV 替换 PORT (F)  -P, --ftp-port ADR 使用指定 PORT 及地址替换 PASV (F)  --ftp-skip-pasv-ip 跳过 PASV 的IP地址 (F)  --ftp-pret 在 PASV 之前发送 PRET (drftpd) (F)  --ftp-ssl-ccc 在认证之后发送 CCC (F)  --ftp-ssl-ccc-mode ACTIVE/PASSIVE 设置 CCC 模式 (F)  --ftp-ssl-control ftp 登录时需要 SSL/TLS (F)  -G, --get 使用 HTTP GET 方法发送 -d 数据 (H)  -g, --globoff 禁用的 URL 队列 及范围使用 {} 和 []  -H, --header LINE 要发送到服务端的自定义请求头 (H)  -I, --head 仅显示响应文档头  -h, --help 显示帮助  -0, --http1.0 使用 HTTP 1.0 (H)  --ignore-content-length 忽略 HTTP Content-Length 头  -i, --include 在输出中包含协议头 (H/F)  -k, --insecure 允许连接到 SSL 站点，而不使用证书 (H)  --interface INTERFACE 指定网络接口／地址  -4, --ipv4 将域名解析为 IPv4 地址  -6, --ipv6 将域名解析为 IPv6 地址  -j, --junk-session-cookies 读取文件中但忽略会话cookie (H)  --keepalive-time SECONDS keepalive 包间隔  --key KEY 私钥文件名 (SSL/SSH)  --key-type TYPE 私钥文件类型 (DER/PEM/ENG) (SSL)  --krb LEVEL 启用指定安全级别的 Kerberos (F)  --libcurl FILE 命令的libcurl等价代码  --limit-rate RATE 限制传输速度  -l, --list-only 只列出FTP目录的名称 (F)  --local-port RANGE 强制使用的本地端口号  -L, --location 跟踪重定向 (H)  --location-trusted 类似 --location 并发送验证信息到其它主机 (H)  -M, --manual 显示全手动  --mail-from FROM 从这个地址发送邮件  --mail-rcpt TO 发送邮件到这个接收人(s)  --mail-auth AUTH 原始电子邮件的起始地址  --max-filesize BYTES 下载的最大文件大小 (H/F)  --max-redirs NUM 最大重定向数 (H)  -m, --max-time SECONDS 允许的最多传输时间  --metalink 处理指定的URL上的XML文件  --negotiate 使用 HTTP Negotiate 认证 (H)  -n, --netrc 必须从 .netrc 文件读取用户名和密码  --netrc-optional 使用 .netrc 或 URL; 将重写 -n 参数  --netrc-file FILE 设置要使用的 netrc 文件名  -N, --no-buffer 禁用输出流的缓存  --no-keepalive 禁用 connection 的 keepalive  --no-sessionid 禁止重复使用 SSL session-ID (SSL)  --noproxy 不使用代理的主机列表  --ntlm 使用 HTTP NTLM 认证 (H)  -o, --output FILE 将输出写入文件，而非 stdout  --pass PASS 传递给私钥的短语 (SSL/SSH)  --post301 在 301 重定向后不要切换为 GET 请求 (H)  --post302 在 302 重定向后不要切换为 GET 请求 (H)  --post303 在 303 重定向后不要切换为 GET 请求 (H)  -#, --progress-bar 以进度条显示传输进度  --proto PROTOCOLS 启用/禁用 指定的协议  --proto-redir PROTOCOLS 在重定向上 启用/禁用 指定的协议  -x, --proxy [PROTOCOL://]HOST[:PORT] 在指定的端口上使用代理  --proxy-anyauth 在代理上使用 "any" 认证方法 (H)  --proxy-basic 在代理上使用 Basic 认证 (H)  --proxy-digest 在代理上使用 Digest 认证 (H)  --proxy-negotiate 在代理上使用 Negotiate 认证 (H)  --proxy-ntlm 在代理上使用 NTLM 认证 (H)  -U, --proxy-user USER[:PASSWORD] 代理用户名及密码  --proxy1.0 HOST[:PORT] 在指定的端口上使用 HTTP/1.0 代理  -p, --proxytunnel 使用HTTP代理 (用于 CONNECT)  --pubkey KEY 公钥文件名 (SSH)  -Q, --quote CMD 在传输开始前向服务器发送命令 (F/SFTP)  --random-file FILE 读取随机数据的文件 (SSL)  -r, --range RANGE 仅检索范围内的字节  --raw 使用原始HTTP传输，而不使用编码 (H)  -e, --referer Referer URL (H)  -J, --remote-header-name 从远程文件读取头信息 (H)  -O, --remote-name 将输出写入远程文件  --remote-name-all 使用所有URL的远程文件名  -R, --remote-time 将远程文件的时间设置在本地输出上  -X, --request COMMAND 使用指定的请求命令  --resolve HOST:PORT:ADDRESS 将 HOST:PORT 强制解析到 ADDRESS  --retry NUM 出现问题时的重试次数  --retry-delay SECONDS 重试时的延时时长  --retry-max-time SECONDS 仅在指定时间段内重试  -S, --show-error 显示错误. 在选项 -s 中，当 curl 出现错误时将显示  -s, --silent Silent模式。不输出任务内容  --socks4 HOST[:PORT] 在指定的 host + port 上使用 SOCKS4 代理  --socks4a HOST[:PORT] 在指定的 host + port 上使用 SOCKSa 代理  --socks5 HOST[:PORT] 在指定的 host + port 上使用 SOCKS5 代理  --socks5-hostname HOST[:PORT] SOCKS5 代理，指定用户名、密码  --socks5-gssapi-service NAME 为gssapi使用SOCKS5代理服务名称  --socks5-gssapi-nec 与NEC Socks5服务器兼容  -Y, --speed-limit RATE 在指定限速时间之后停止传输  -y, --speed-time SECONDS 指定时间之后触发限速. 默认 30  --ssl 尝试 SSL/TLS (FTP, IMAP, POP3, SMTP)  --ssl-reqd 需要 SSL/TLS (FTP, IMAP, POP3, SMTP)  -2, --sslv2 使用 SSLv2 (SSL)  -3, --sslv3 使用 SSLv3 (SSL)  --ssl-allow-beast 允许的安全漏洞，提高互操作性(SSL)  --stderr FILE 重定向 stderr 的文件位置. - means stdout  --tcp-nodelay 使用 TCP\_NODELAY 选项  -t, --telnet-option OPT=VAL 设置 telnet 选项  --tftp-blksize VALUE 设备 TFTP BLKSIZE 选项 (必须 >512)  -z, --time-cond TIME 基于时间条件的传输  -1, --tlsv1 使用 => TLSv1 (SSL)  --tlsv1.0 使用 TLSv1.0 (SSL)  --tlsv1.1 使用 TLSv1.1 (SSL)  --tlsv1.2 使用 TLSv1.2 (SSL)  --trace FILE 将 debug 信息写入指定的文件  --trace-ascii FILE 类似 --trace 但使用16进度输出  --trace-time 向 trace/verbose 输出添加时间戳  --tr-encoding 请求压缩传输编码 (H)  -T, --upload-file FILE 将文件传输（上传）到指定位置  --url URL 指定所使用的 URL  -B, --use-ascii 使用 ASCII/text 传输  -u, --user USER[:PASSWORD] 指定服务器认证用户名、密码  --tlsuser USER TLS 用户名  --tlspassword STRING TLS 密码  --tlsauthtype STRING TLS 认证类型 (默认 SRP)  --unix-socket FILE 通过这个 UNIX socket 域连接  -A, --user-agent STRING 要发送到服务器的 User-Agent (H)  -v, --verbose 显示详细操作信息  -V, --version 显示版本号并退出  -w, --write-out FORMAT 完成后输出什么  --xattr 将元数据存储在扩展文件属性中  -q .curlrc 如果作为第一个参数无效 |

|  |
| --- |
| Hdfs上新建文件   1. curl -i -X PUT "http://172.26.41.32:32055/webhdfs/v1/tmp/datatom/rensheju\_dsjzx\_zwy\_syjlq\_xxb.txt?op=CREATE&guardian\_access\_token=6mKYTUMtterzxlYyMpv4-CEIJAZ5.TDH"   上传本地文件至hdfs上  (2)curl -i -C - -X PUT -T /opt/datatom/dana\_api/data/rensheju/dsjzx\_zwy\_syjlq\_xxb/dsjzx\_zwy\_syjlq\_xxb\_split/dsjzx\_zwy\_syjlq\_xxb\_\_cd93321d\_712f\_4908\_b29d\_48978a4939f90 "http://172.17.148.189:31734/webhdfs/v1/tmp/datatom/rensheju\_dsjzx\_zwy\_syjlq\_xxb.txt?op=CREATE&data=true&guardian\_access\_token=EAucQR4SuHCcVXt10VeG-M8IUIMT.TDH" -H "Content-Type:application/octet-stream"  **加载数据入库**   1. 往orc表中插入数据   load **DATA** inpath '/tmp/datatoms/hfhmy.txt' **INTO** **TABLE** `rexian\_txt`.`ods\_12345\_hfhmydxxys`   1. 从hdfs上insert数据   **INSERT** **INTO** `rexian\_txt`.`ods\_12345\_hfhmydxxys\_tmp`  **SELECT** \* **FROM** `rexian\_txt`.`ods\_12345\_hfhmydxxys` |

|  |
| --- |
| **分区表插入数据示范**  SET hive.exec.dynamic.partition=true; #开启动态分区插入数据的开关  insert into gonganju\_orc\_ods.base\_sjkk\_gcjl\_ecjgh\_df (month\_partition) select ...,substring(rksj,1,7) from gonganju\_txt\_ods.base\_sjkk\_gcjl\_ecjgh\_df;  说明：  month\_partition : 分区表的分区字段  ... : 为txt表的全字段  substring(rksj,1,7) : 作为分区字段的业务字段的函数 |

## 星环建表

### Text表

|  |
| --- |
| **CREATE** **TABLE** rensheju\_txt.zwy\_syjlq\_xxb  (  ZJHM **VARCHAR2**(18) **NOT** **NULL**,  ZJLX **CHAR**(2) **NOT** **NULL**,  XM **VARCHAR2**(30) **NOT** **NULL**,  GJ **VARCHAR2**(80) **NOT** **NULL**,  BTLX **VARCHAR2**(80) **NOT** **NULL**,  NY **CHAR**(6) **NOT** **NULL**,  LQYS **NUMBER**(3,0) **NOT** **NULL**,  ZJE **NUMBER**(16,2) **NOT** **NULL**,  JHPT\_UPDATE\_TIME **TIMESTAMP**(0) **DEFAULT** SYSDATE,  JHPT\_UPDATE\_FLAG **NUMBER**(1,0) **DEFAULT** 0,  JHPT\_DELETE **VARCHAR2**(6),  YW\_UPDATE\_TIME **TIMESTAMP**(6))  **ROW** FORMAT DELIMITED FIELDS TERMINATED **BY** '\u0001'  lines terminated by '!@#';（如果是文件需要指定行分隔符） |

### Orc事务表

|  |
| --- |
| **CREATE** **TABLE** rensheju\_orc.zwy\_syjlq\_xxb  (  ZJHM **VARCHAR2**(18) **NOT** **NULL**,  ZJLX **CHAR**(2) **NOT** **NULL**,  XM **VARCHAR2**(30) **NOT** **NULL**,  GJ **VARCHAR2**(80) **NOT** **NULL**,  BTLX **VARCHAR2**(80) **NOT** **NULL**,  NY **CHAR**(6) **NOT** **NULL**,  LQYS **NUMBER**(3,0) **NOT** **NULL**,  ZJE **NUMBER**(16,2) **NOT** **NULL**,  JHPT\_UPDATE\_TIME **TIMESTAMP**(0) **DEFAULT** SYSDATE,  JHPT\_UPDATE\_FLAG **NUMBER**(1,0) **DEFAULT** 0,  JHPT\_DELETE **VARCHAR2**(6),  YW\_UPDATE\_TIME **TIMESTAMP**(6))  CLUSTERED **BY** (logs\_id) **INTO** 4 BUCKETS  STORED **AS** ORC  TBLPROPERTIES ("transactional"="true"); |

### Csv表

|  |
| --- |
| create table yibaoju\_txt.MATBUYINFO\_csv(  ID string,  HCTBDM string,  NY string,  YYBM string,  QYBM string,  CGSL string,  CGJE string,  JHPT\_UPDATE\_TIME string,  JHPT\_DELETED string,  ADDTIME string  )  stored as csvfile; |

### 设置多字段分隔符建表

|  |
| --- |
| CREATE TABLE shuiwuju\_txt.SB\_SDS\_JMCZ\_14ND\_JMSDSYHMX  (  sbuuid string comment '申报UUID',  pzxh decimal(20) comment '凭证序号',  gjghzdjcdlsjqy decimal(18, 2) comment '国家规划布局内重点集成电路设计企业',  xkxyldbwmdjcdlqy decimal(18, 2) comment '集成电路线宽小于0.8微米（含）的集成电路生产企业',  cgbsyjcdlscqy decimal(18, 2) comment '投资额超过80亿元人民币的集成电路生产企业',  xbjcdlsjqy decimal(18, 2) comment '新办集成电路设计企业',  fhtjrjqy decimal(18, 2) comment '符合条件的软件企业',  gjghzdrjqy decimal(18, 2) comment '国家规划布局内重点软件企业',  jcdlfzcsqy decimal(18, 2) comment '集成电路封装、测试企业',  jcdlgjzyclscsbqy decimal(18, 2) comment '集成电路关键专用材料生产企业或集成电路专用设备生产企业',  xkxyldewwmdjcdlqy decimal(18, 2) comment '集成电路线宽小于0.25微米的集成电路生产企业',  lrrq timestamp comment '录入日期',  jhpt\_update\_time timestamp comment '时间戳',  jhpt\_delete int comment '删除标志'  )ROW FORMAT SERDE 'org.apache.hadoop.hive.contrib.serde2.MultiDelimitSerDe' WITH SERDEPROPERTIES ('input.delimited'='||') stored as textfile; |