程序说明

文档控制说明

**修订历史**

| **版本号** |  | **修订人** | **变更内容** | **备注** |
| --- | --- | --- | --- | --- |
| V0.1 | 2012-11-13 | 崔雪征 | 创建文档 |  |
|  |  |  |  |  |

# 概要

## 范围

此处的程序是指为完成ETL中业务逻辑，或者日常任务处理而加的小的程序，不包括调度等大程序。

# 明细

## 系统辅助

### hive\_daemon.sh

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 名称 | hive\_daemon.sh | | | | | | | |
| 描述 | HiveServer守护进程 | | | | | | | |
| 输入参数 | | | | | | | | |
| 名称 | | 类型 | | | 缺省值 | | | 描述 |
|  | |  | | |  | | |  |
| 输出参数 | | | | | | | | |
| 类型 | | | | 缺省值 | | | 描述 | |
|  | | | |  | | |  | |
| 处理逻辑 | | | | | | | | |
| 每隔一段时间检测HiveServer进程是否存在，如果不存在，则启动此进程 | | | | | | | | |
|  | | | | | | | | |
| 配置文件或表 | | | | | | | | |
| 文件或表名 | | | 数据库 | | | 备注 | | |
|  | | |  | | |  | | |
|  | | |  | | |  | | |
| 备注 | | | | | | | | |
|  | | | | | | | | |

### comm\_func.sh

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 名称 | comm\_func.sh | | | | | | | |
| 描述 | 通用函数及常量 | | | | | | | |
| 输入参数 | | | | | | | | |
| 名称 | | 类型 | | | 缺省值 | | | 描述 |
|  | |  | | |  | | |  |
| 输出参数 | | | | | | | | |
| 类型 | | | | 缺省值 | | | 描述 | |
|  | | | |  | | |  | |
| 处理逻辑 | | | | | | | | |
| SHELL中通用的函数及变量。 | | | | | | | | |
|  | | | | | | | | |
| 配置文件或表 | | | | | | | | |
| 文件或表名 | | | 数据库 | | | 备注 | | |
|  | | |  | | |  | | |
|  | | |  | | |  | | |
| 备注 | | | | | | | | |
|  | | | | | | | | |

## ETL处理

### mkdir.sh

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 名称 | mkdir.sh | | | | | | | |
| 描述 | 创建目录 | | | | | | | |
| 输入参数 | | | | | | | | |
| 名称 | | 类型 | | | 缺省值 | | | 描述 |
| day\_id | |  | | |  | | | 日期ID |
| 输出参数 | | | | | | | | |
| 类型 | | | | 缺省值 | | | 描述 | |
|  | | | |  | | |  | |
| 处理逻辑 | | | | | | | | |
| 1、遍历配置文件 | | | | | | | | |
| 1.1 把未来几天文件内容中的yyyymmdd用具体日期替换，如果没有此目录则创建 | | | | | | | | |
| 配置文件或表 | | | | | | | | |
| 文件或表名 | | | 数据库 | | | 备注 | | |
| dir.conf | | |  | | | 只有一列，带有#yyyymmdd#日期参数的文件目录 | | |
|  | | |  | | |  | | |
| 备注 | | | | | | | | |
| 目前只实现了按天建目录。 | | | | | | | | |

### exp\_inf.sh

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 名称 | exp\_inf.sh | | | | | | | |
| 描述 | 创建目录 | | | | | | | |
| 输入参数 | | | | | | | | |
| 名称 | | 类型 | | | 缺省值 | | | 描述 |
| day\_id | |  | | |  | | | 日期ID |
| ctl\_file | |  | | |  | | | 控制文件名称 |
| 输出参数 | | | | | | | | |
| 类型 | | | | 缺省值 | | | 描述 | |
|  | | | |  | | |  | |
| 处理逻辑 | | | | | | | | |
| 1、根据传入的控制文件查找导出文件的路径和目标路径； | | | | | | | | |
| 2、将hive导出的文件合并成一个目标文件； | | | | | | | | |
| 3、生成控制文件内容 | | | | | | | | |
| 配置文件或表 | | | | | | | | |
| 文件或表名 | | | 数据库 | | | 备注 | | |
| exp\_inf.conf | | |  | | | 三列，分别是控制文件、Hive导出文件路径、接口文件路径 | | |
|  | | |  | | |  | | |
| 备注 | | | | | | | | |
| 简单实现，复杂的处理日后再加 | | | | | | | | |