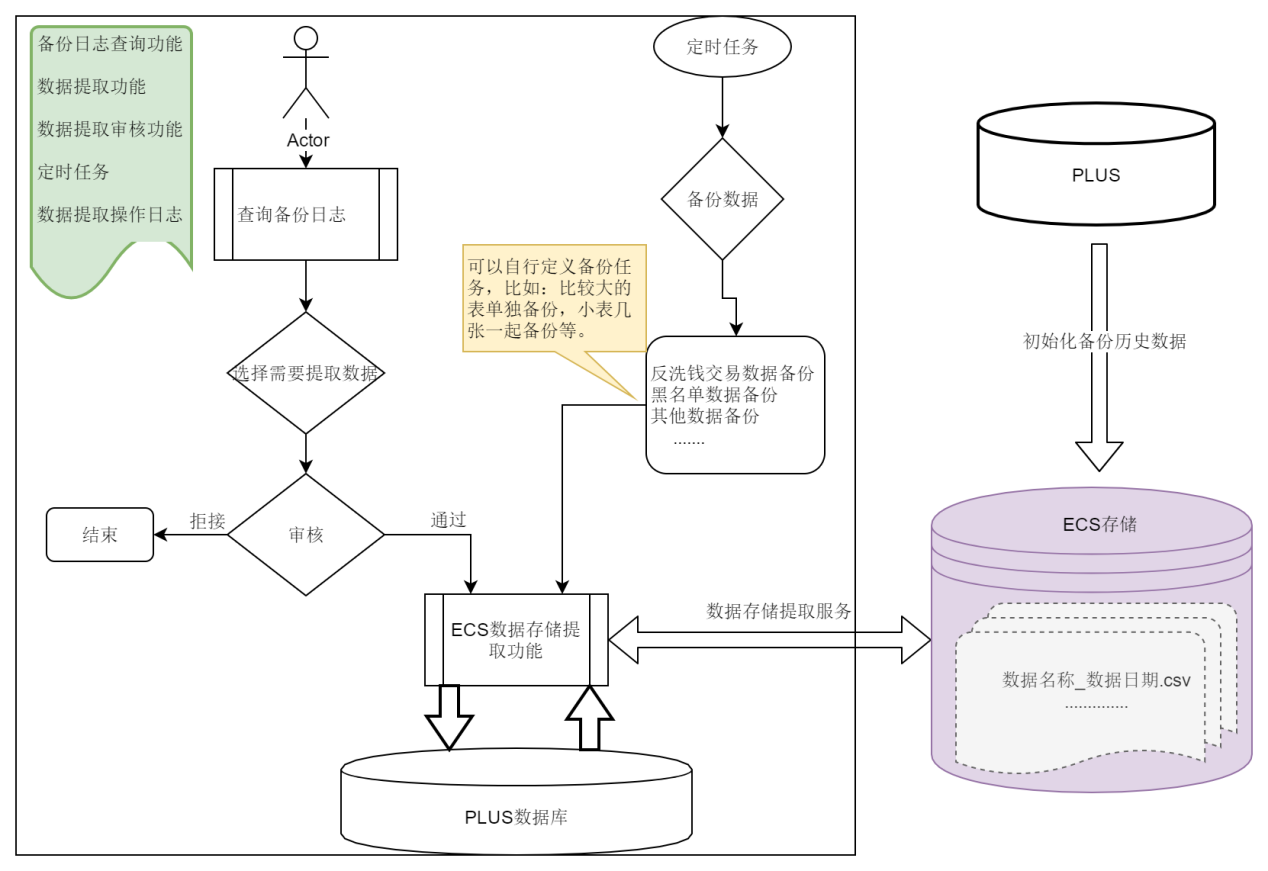
PLUS数据平台数据备份&提取方案

|  |  |  |
| --- | --- | --- |
| 撰写人 | 撰写时间 | 版本号 |
| 武利兵 | 2021-03-01 | 0.1 |

1. 方案说明

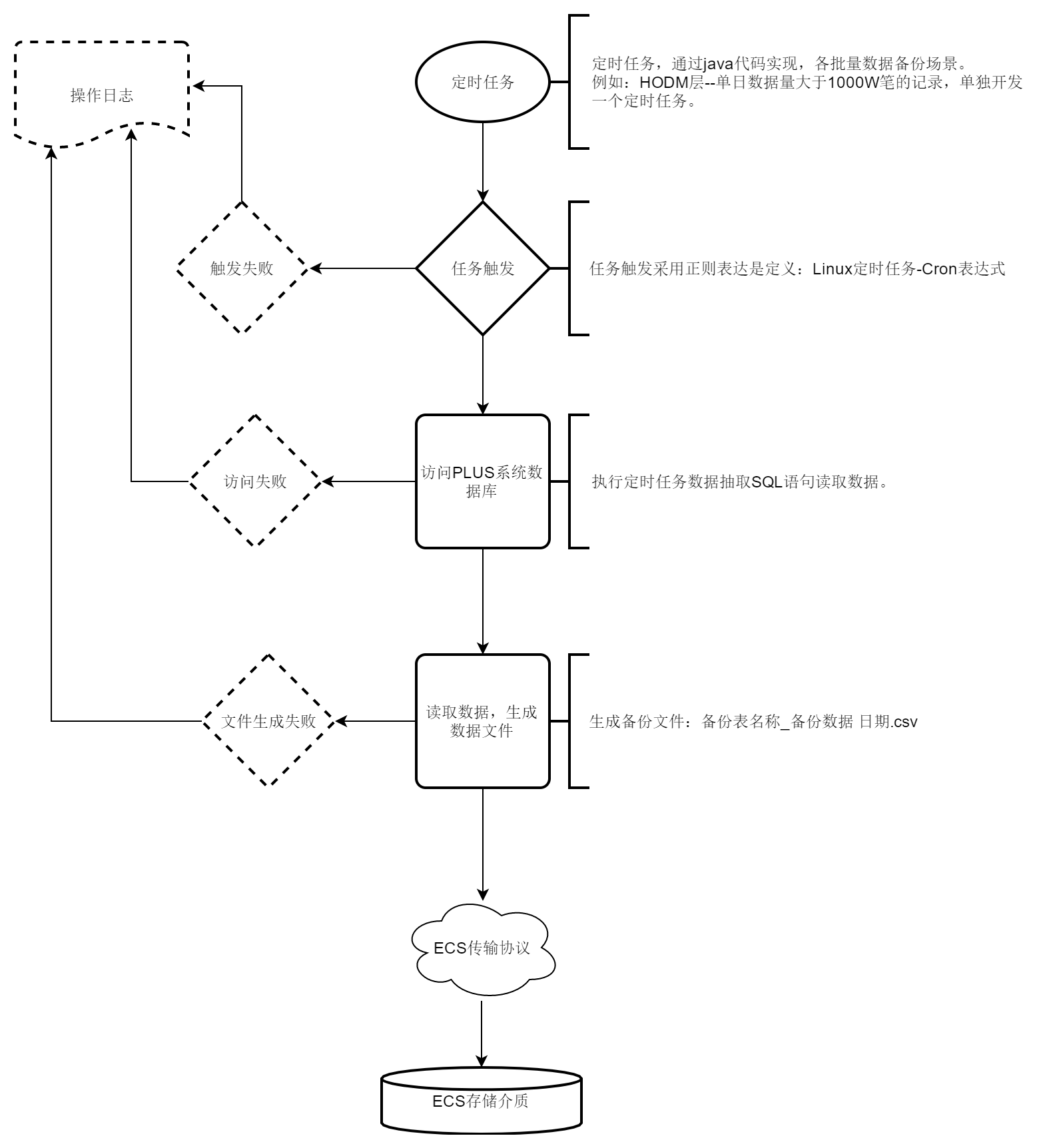
汇丰PLUS系统，2017年上线至今，积累了4年的数据量，同时为下游EAST，MBT,CRMS提供数据支持服务。考虑数据存储成本的逐年增加，数据量过大影响每日批量运行效率、数据查询效率，以及下游应用系统对数据的访问和其特殊的业务需要等情形。设计本方案来缓解PLUS应用系统在使用数据时，批量运行效率、查询效率上的问题，同时解决下游应用系统在使用数据时，对ARK数据平台的使用依赖。

方案流程图：



* **定时任务：**数据备份任务，设定数据表的备份范围，备份数据的数据量区间，任务运行时间等。
* **数据提取：**通过此功能提取已存储在ECS中的备份数据。
* **备份日志查询：**记录已备份数据信息，备份文件名称，备份日期，备份数据范围等信息。
* **数据提取审核：**规范数据安全，对提取数据工作进行管理。
* **数据提取操作日志：**记录提取操作时的工作日志，提取人，提取时间，提取数据表清单，提取结果等信息。
* **ECS存储：**存储环境，已文件形式存储PLUS系统中备份的数据。
* **PLUS数据库：**oracle 数据库。

1. 数据存储
   1. 存储流程

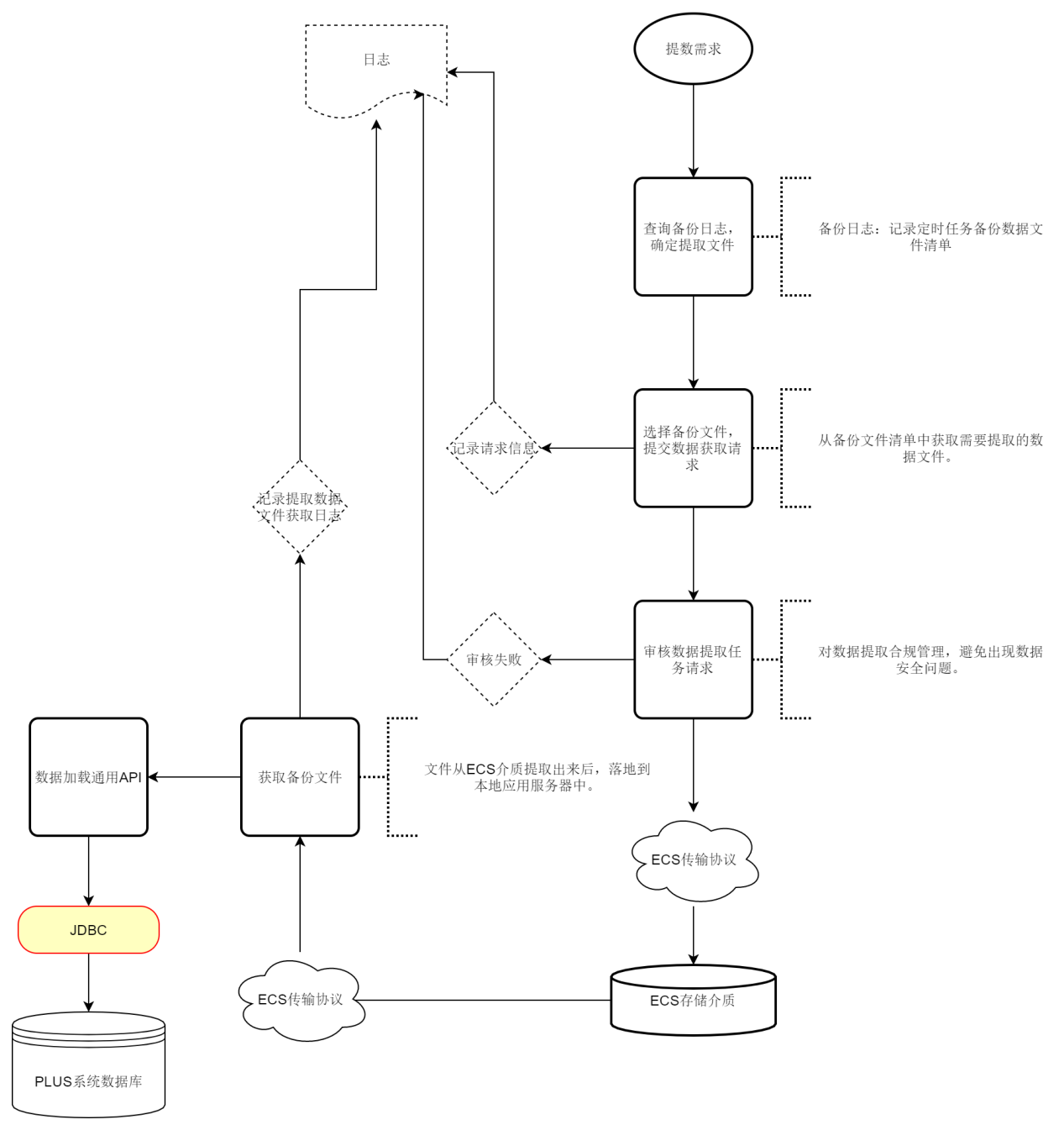


* 1. 规则标准

备份文件格式定义：

* 文件名格式：文件名称+”\_”+数据日期+”.csv”
* 文件分隔符：“||”双竖线
* 文件编码格式：UTF-8
* 文件备份规则：表当日记录大于1000万笔的单独备份，表记录小于1000万大于100万的记录，10张表合并备份，表记录小于100的，30张表合并备份，合并压缩成zip文件。命名规则同文件名格式规则。
* 数据清理规则：当日备份完成，并获得ECS存储回执。第二日执行清理脚本。清理T-1日备份数据。

1. 数据提取

 **数据加载通用API：**对提取到应用服务器上的数据文件，进行数据加载处理。可以进行简单的加载数据库名称、用户、密码参数输入，加载文件选择，加载表映射，加载字段选择，加载字段去空处理、格式转换等功能。