# 数据平台--衍生品对接方案

1. **系统需求**
   1. **接口列表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **需求方** | **需求接口名称** | **结构化标志** | **传输文件格式** | **接口编码** | **说明** |
|  | 机构监管 | 主协议 | 结构化 | TXT | A1001 |  |
|  | 机构监管 | 补充协议 | 结构化 | TXT | A1002 |  |
|  | 机构监管 | 期权交易确认书 | 结构化 | TXT | A1003 |  |
|  | 机构监管 | 互换交易确认书 | 结构化 | TXT | A1004 |  |
|  | 机构监管 | 互换及期权终止 | 结构化 | TXT | A1005 |  |
|  | 机构监管 | 履约保障 | 结构化 | TXT | A1006 |  |
|  | 机构监管 | 重大事项报告 | 结构化 | TXT | A1007 |  |
|  | 机构监管 | 其他事项报告 | 结构化 | TXT | A1008 |  |
|  | 机构监管 | 定期报告-SAC模板 | 结构化 | TXT | A1009 |  |
|  | 机构监管 | 定期报告-NAFMII模板 | 结构化 | TXT | A1010 |  |
|  | 机构监管 | 定期报告-ISDA模板 | 结构化 | TXT | A1011 |  |

* 1. **文件接口规范**
     1. **文件类型**

文件采集以接口为基本报送单位，涉及的文件类型有以下两种:

结构化数据文件：指最终报送的结构化数据文件，采用TXT格式。

非结构化数据文件：指最终报送的数据文件为PDF等非结构化格式的文件。

**数据文件传输过程中，首先写入以tmp为扩展名的临时文件，待写入完毕后，再更名为接口中约定扩展名（TXT或PDF等）**。

* + 1. **结构化数据**

文件采集以接口为基本报送单位，最终报送的文件类型均是结构化数据文件，传输文件格式为TXT格式。该文件主要用于采集结构化业务数据。参见3.1数据接口列表。

**数据文件传输过程中，首先写入以tmp为扩展名的临时文件，待写入完毕后，再更名为接口中约定扩展名（TXT）**。

* + 1. **非结构化数据**

对于非结构化数据，需要统一放到ATTACHMENT目录下进行传输。该文件主要用于采集非结构化业务数据，如：产品合同、募集说明书等。

* + 1. **命名规则**

数据文件命名规则如下：

OTC\_[文件创建人代码]\_[文件接收人代码]\_[接口标识]\_[数据日期]\_[文件序号]\_[操作标识].[文件扩展名]

* 文件名中所有字母均为大写。
* 文件创建人代码:6位字符，参与人会员编码，使用场外证券业务报告系统为机构分配的会员编码。
* 文件接收人代码:6位字符，参与人会员编码，使用场外证券业务报告系统为机构分配的会员编码，如果接收方为，场外证券业务报告系统，统一默认填写000899。
* 接口标识：接口规范中的接口编号，如：A0001。
* 日期为8位字符，数字日期，格式为YYYYMMDD
* 文件序号为4位，每日从“0001”开始，最大支持9999。
* 操作标识：该字段1位字符，具体见下表：

|  |  |  |
| --- | --- | --- |
| **操作类型** | **标识符** | **说明** |
| 报送信息文件 | A | 报送时使用 |
| 应答文件 | R | 返回时使用 |

* 文件扩展名：数据文件的扩展名，采用TXT 格式。

**示例：**

OTC\_131101\_000899\_A0001\_20171211\_0001\_A.TXT

* 1. **接口交互模式**
     1. **报送信息文件**

报送信息文件是机构上报的衍生品业务信息，格式为结构化数据文件、非结构化数据文件。为满足数据交互中对应关系， 以下是为所有数据接口额外补充的接口辅助字段，字段顺序排在数据接口字段前，具体信息如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **字段标识** | **字段中文名** | **数据类型** | **是否必填** | **说明** |
| 1 | ExecID | 报送记录号 | C（28） | 是 | 文件创建人代码（6位）+文件接收人代码（6位）+报送日期（8位）+当日顺序编号（8位） |

以主协议（A1001）报送为例，组织结构参见如下：



同时上报的还需包含"ATTACHMENT"文件夹，以及对于的非结构化附件文件。



* + 1. **通用应答文件**

通用应答文件是场外证券业务报告系统在接收到证券公司、证券投资基金公司、期货公司、证券投资咨询机构、私募基金管理人、证券资信评级、证券增信机构报备的报文后，用此文件向报送机构回复“报送信息文件”。应答文件格式为结构化数据文件，具体返回信息如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **字段标识** | **字段中文名** | **数据类型** | **是否必填** | **说明** |
| 1 | AnsID | 应答记录号 | C(28) | 是 | 文件创建人代码（6位）+文件接收人代码（6位）+报送日期（8位）+当日顺序编号（8位） |
| 2 | ExecID | 原报送记录号 | C(28) | 否 | 对应原报送记录 |
| 3 | ExecType | 处理结果 | C(1) | 是 | 执行类型代码（Y =数据处理成功，N = 数据处理失败） |
| 4 | TransactTime | 处理时间 | C(20) | 是 | 执行的创建时间  YYYMMDD HH:MM:SS |
| 5 | AnsDetail | 处理信息 | V(4000) | 否 | 错误码及错误信息参见附录 |

主协议报送成功应答示例：



主协议报送失败应答示例：



* 1. **加密方式**

为提高数据安全性，报送业务数据前，需对业务数据进行加密处理，然后再进行报送。加密数据包括结构化、非结构化数据；加密工具是由场外证券业务报告库系统发的证书。

* 1. **传输方式**

该文件接口规范支持通过证联网、深圳通方式进行传输。

1. **系统设计**
   1. **数据平台总概**
      1. **数据报送场景**
         1. **日常报送**

日常报送是指数据源单位按照规范规定，进行日常周期性数据报送的场景。

日常报送场景下，数据源单位应按照接口数据模型，生成该报送日的数据文件，在约定报送日时间段(下面有描述)报送至数据平台，若遇节假日，报送日顺延至其后的第一个交易日。

数据平台只在交易日的上午9:00至下午3：00运行处理文件上报操作。

* + - 1. **异常情况报送**

异常情况报送是指数据源单位未能按照接口规范的规定，将数据正常报送至数据平台的场景。

* + - * 1. **数据重报**

当数据源单位报送至数据平台的数据接口文件未满足相应的数据质量约束条件时，需进行数据重报。数据重报时，数据源单位应产生新的数据接口文件，新的数据接口文件序号应为上一次报送数据文件的文件序号加1。

**示例：**

OTC\_131101\_000899\_A0001\_20171211\_0001\_A.TXT

OTC\_131101\_000899\_A0001\_20171211\_0002\_A.TXT

* + 1. **库表设计**

1. **处理文件表** **handle\_file\_tab**

存储压缩文件及文件名等信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| file\_name | varchar | 100 | 文件名 |
| hash\_file\_entity | varchar | 50 | 文件内容哈希值 |
| file\_creater\_code | varchar | 10 | 文件创建人代码 |
| file\_receiver\_code | varchar | 10 | 文件接收人代码 |
| interface\_id | varchar | 10 | 接口标识 |
| submission\_date | varchar | 10 | 报送日期 |
| file\_num | varchar | 10 | 文件序号 |
| operation\_id | varchar | 2 | 操作标识 |
| status | int | 2 | 状态 |
| create\_date | datetime | 0 | 创建时间 |
| update\_date | datetime | 0 | 修改时间 |

1. **处理明细表 handle\_detail\_tab**

存储解析文件内容等信息

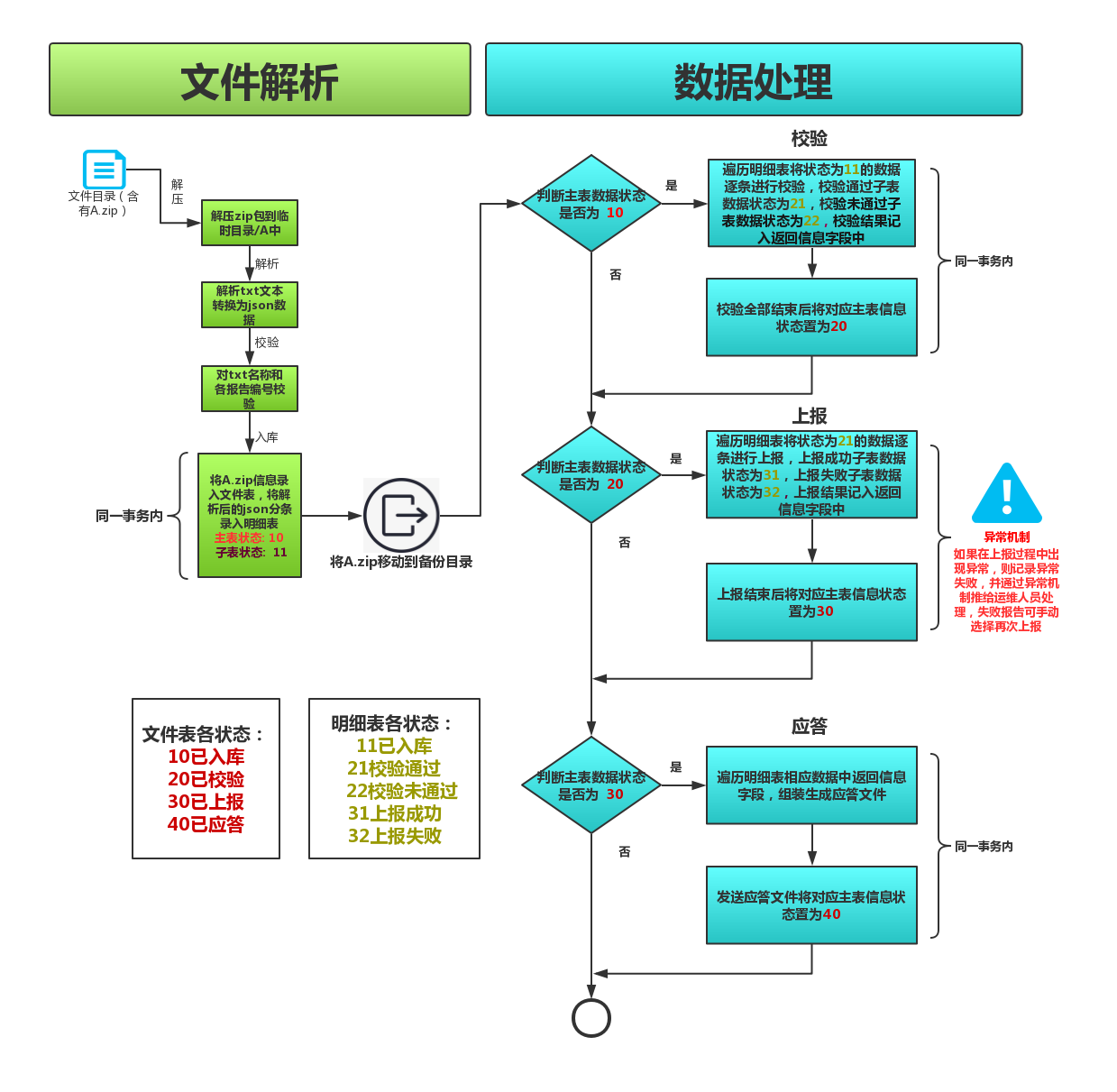
|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| file\_id | int | 20 | 处理文件表id |
| file\_creater\_code | varchar | 10 | 文件创建人代码 |
| file\_receiver\_code | varchar | 10 | 文件接收人代码 |
| interface\_id | varchar | 10 | 接口标识 |
| submission\_date | varchar | 10 | 报送日期 |
| file\_num | varchar | 10 | 文件序号 |
| operation\_id | varchar | 2 | 操作标识 |
| submission\_record\_id | varchar | 50 | 报送记录号 |
| data\_json | longtext | 0 | json数据 |
| status | int | 2 | 状态 |
| response\_msg | varchar | 2000 | 应答信息 |
| create\_date | datetime | 0 | 创建时间 |
| update\_date | datetime | 0 | 修改时间 |

1. **系统日志表 sys\_logging\_tab**

存储解析文件内容等信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| logging\_type | int | 2 | 日志类型  10:自己  11:校验库  12:报告库 |
| error\_code | varchar | 10 | 错误码 |
| error\_msg | varchar | 2000 | 错误信息 |
| create\_date | datetime | 0 | 创建时间 |

* + 1. **处理文件流程图**



* + 1. **处理文件描述**
* 扫描指定目录文件（zip）代号为A.ZIP，解压A.zip文件到临时目录/A中，获取B.txt文件流得到MD5哈希值(长度32),根据字段hash\_file\_entity查询处理文件表中是否存在数据，若存在则是重复传送且内容重复的文件，解析B.txt文本转化为json数据，生成应答文件(失败原因："该报告已上报过，请勿重复上报")；若不存在则获取B.txt名称信息(文件创建人代码、文件接收人代码、接口标识、数据日期、文件序号、操作标识)，保存录入处理文件表，数据状态为"已入库：10"，解析B.txt文本转化为json数据，根据json数据条取出报送记录号和B.txt名称信息(文件创建人代码、文件接收人代码、接口标识)和状态"上报成功：31"查询处理明细表数据，若不存在则为数据正常报送；若存在则为数据重报；

**数据正常报送**

json数据条录入处理明细表，数据状态为"已入库：11"，保存完毕后，将A.zip移动到备份目录中，以上操作都在一个事务里。

**数据重报**

数据移植进处理明细表，将状态变为"上报成功：32"，失败原因："该报告已上报过，请勿重复上报"；反之则json数据条录入处理明细表，数据状态为"已入库：11"；保存完毕后，将A.zip移动到备份目录中，以上操作都在一个事务里。

* 查询处理文件表数据状态为"已入库：10"的数据，遍历处理明细表数据状态为"已入库：11"进行校验，若校验通过，处理明细表数据状态置为"校验通过：21"，若校验未通过，处理明细表数据状态置为"校验未通过：22"并记下校验未通过原因，遍历完毕则处理文件表数据状态置为"已校验：20"，以上操作都在一个事务里。
* 查询处理文件表数据状态为"已校验：20"的数据，遍历处理明细表数据状态为"校验通过：21"进行上报，若上报成功，处理文件表数据状态置为"上报成功：31"，若上报失败，处理明细表数据状态置为"上报失败：32"并记下上报失败原因，遍历完毕则处理文件表数据状态置为"已上报：30"，以上操作没有事务，出现异常情况通过异常机制告知运营人员，记录日志并终止程序。
* 查询处理文件表数据状态为"已上报：30"的数据，遍历处理明细表数据生成应答文件，遍历完毕则处理文件表数据状态置为"已应答：40"，以上操作都在一个事务里。
* 现考虑两种方案：

将以上4步骤都放在1个定时任务里处理，优点便于维护，缺点执行效率慢。

将以上4步骤都放在4个定时任务里处理，优点执行效率快，缺点不便于维护。

* 1. **数据解析**

过程三：解析并校验文件

服务端程序会在某一时间对上传的数据文件进行解析，使用java IO中FileInputStream和BufferedReader获取文件流，在BufferedReader中有readLine()方法，使用该方法完成按行读取数据文件内容。

解析分为以下四种情况：

* + 1. **非业务数据解析**

读取文件名称，将名称以“\_”进行分隔，获取数组可得到报送方、接收方、业务数据类型。如文件OTC\_131101\_000899\_A0001\_20171211\_0001\_A.TXT

解析后可获得

发送方编号 131101

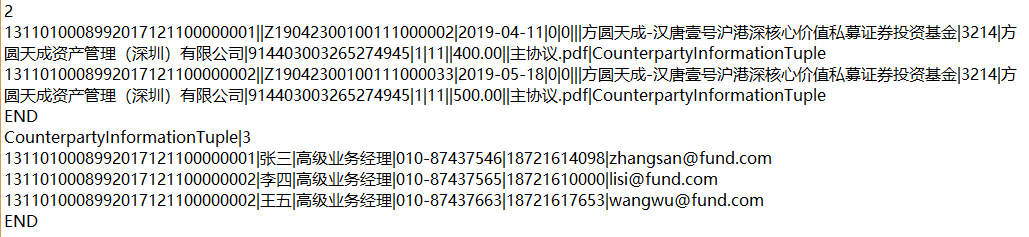
接收方编号 000899

业务类型 A0001

* + 1. **业务数据解析**

将结构化数据文件内容读取为文件流，然后按行读取

例如文件内容如下：



第一行为总数据条数，当最后解析完毕后需要校验解析总条数是否等于该数据；

第二行开始为业务数据，将得到的字符串用“|”分隔，获取数组数据，将此数组按照规定的顺序匹配对应的业务字段生成java实例，同时使用校验库规则，将该数据进行校验（具体处理见校验库设计），如若检验不通过生成校验结果文件。

第三行 处理同上… …

直到读取数据内容为“END”字符串，标识此业务数据主列表信息读取完毕。

* + 1. **非结构化数据处理**

当解析业务数据遇到文件类型的数据时，需要将该文件名称、存储目录、业务数据ID等数据获取生成java数据对象；

* + 1. **子列表数据解析**

当解析数据遇到子列表标识不为空的情况时，在读取到“END”时，继续按行读取数据

第一行数据，将得到的字符串按“|”进行分割，得到数组，数组中的第一个元素为对应业务字段的标识，第二个元素为子列表中应该解析的总条数，用于解析完毕后的检验；

第二行开始为业务数据，将得到的字符串用“|”分隔，获取数组数据，将此数组按照规定的顺序匹配对应的业务字段生成java实例，同时使用校验库规则，将该数据进行校验（具体处理见校验库设计），如若检验不通过生成校验结果文件；

第三行 处理同上… …

直到读取数据内容为“END”字符串，标识此业务子列表数据信息读取完毕。

* 1. **数据存储**

过程4：将过程3中解析并校验通过的业务数据存储到数据库。

* 1. **数据校验**
     1. **校验库设计**

设计该校验库使所有通过数据传输的报告能统一校验，所有校验规则和字段内容写在数据库中，能够实时更新校验内容和字段而不用更新代码和程序，方便后期变更和管理。减少对程序的修改。

设计原则：

1. 入参数据统一格式为json数据，统一解析数据。
2. 将报告库和其他系统报告按照逐条报送校验（性能考虑）
   * 1. **校验库表设计**
3. 模块表

用于区分各模块类型和报告类型及信息

1. 字段表

存储各字段信息，类型，长度，校验规则码等信息

1. 校验规则表

存储各校验规则信息，校验码，校验错误提示语，校验类型

**1模块表**

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| type\_id | varchar | 50 | 主键 |
| type\_code\_interface | varchar | 50 | 报告编码 |
| t\_template\_type | varchar | 50 | 模块类型 |
| type\_name | varchar | 100 | 中文名称 |
| remark | varchar | 2000 | 备注 |

**2字段表**

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| Id | Int | 20 | 主键 |
| template\_type | varchar | 50 | 模块类型 |
| template\_name | varchar | 200 | 字段名称 |
| template\_validation | varchar | 50 | 校验编码 |
| D\_field | varchar | 50 | 数据库字段 |
| D\_field\_type | varchar | 50 | 字段类型 |
| D\_table\_name | varchar | 100 | 字段所在表名 |
| status | varchar | 20 | 字段状态 |
| dTupleUpName | varchar | 50 | 上级tuple数据库字段名称 |
| tControlType | varchar | 50 | 控件类型 |
| remark | varchar | 2000 | 备注 |

**3校验规则表**

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| R\_ID | Int | 10 | 主键 |
| R\_key | varchar | 50 | 校验标识 |
| R\_name | varchar | 255 | 校验名称 |
| R\_type | varchar | 50 | 校验类型 |
| R\_CONTENT | varchar | 2000 | 校验规则 |
| remark | varchar | 2000 | 校验错误返回信息提示语 |

* + 1. **校验逻辑**

数据加载后，先从报告模版表中取出相应的字段名，校验规则编号，在通过校验规则编号去规则表中取出相应规则和规则类型，遍历每个入参数据根据名称找到对应的规则类型进行校验，（类型分为联动必填，联动控制，联动日期控制等）联动校验使用正则表达式的形式校验，联动的数据必须都在同一数据报告里。也可通过代码自己建立新的校验类型及规则如：

添加类型 match 添加规则ZQS+HHS=ZS(期权数+互换数=总数) 在代码中写好match类型校验代码 在模板表ZS字段上添加该校验，所填写的ZS就必须是两数之和，其他校验也可类似添加。

* + 1. **校验库扩展及后期管理**

**扩展：**

当校验业务有所变更需要改动时时，只需改动数据库配置，在相应字段加入校验规则即可，不用改动程序。如有新的特殊校验类型就要在代码里写好如何处理逻辑及类型，然后在校验规则表中加入该校验，最后在相应字段加入新的校验规则码即可，以后再有相同的校验就可直接配置不用再次修改代码。

**管理：**

日常对于校验表和字段表的维护可以写成程序和页面来管理，通过程序对校验库进行增删改查的操作，代替sql脚本的操作，使系统更加安全有效减少误操作。

校验日志记录所有报告的出参及返回信息，使用logback记录日志并输出到指定文件。

**备份与导出：**

每日的详细规则会自动备份在指定目录下，也可通过管理页面导出excel信息。

* + 1. **参数管理**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 输入参数 （JSON格式） | 参数名 | 定义 | 类型 | 必须 | 说明 |
| bizId | 报告业务编号 | String | Y | 报告业务编号 |
| type | 报告类型 | String | Y | 报告类型参见 |
| orgId | 机构ID | String | Y |  |
| userId | 用户ID | String | Y |  |
| data | 业务数据 | String | Y |  |
| 输出参数 （JSON格式） | code | 结果码 | String | Y | 成功|0、 失败|-1; |
| msg | 提示信息 | String | Y | 返回提示信息 |
| bizId | 报告业务编号 | String | Y | 报告业务编号 |

* + 1. **整体流程**

①报告数据通过接口一json格式进入校验后台

②校验程序读取数据库

③加载每个字段基础校验信息，校验类型及长度

④加载每个字段校验规则码，校验对应逻辑规则

⑤校验后返回相应信息

后台程序

需校验的报告数据

获取报告数据，与数据库交互获取校验规则，类型，数据类型，长度

基本校验

校验报告数据的类型和长度

逻辑校验

根据校验类型找到相应的校验代码，读取校验规则进行数据校验

（根据校验类型增多代码也会变动）

数据库

根据报告类型取出数据





* 1. **数据报送**

查询主表数据状态为"未完成：2"的数据，遍历子表数据找到各类报告，组装成报文后调用报告库接口进行上报，其中包括主、补、履各协议和期权与互换的交易确认书，并将附件一并上传。

若上报成功，子表数据状态置为"上报成功：4"，将返回编号记录在返回信息中，若上报失败，子表数据状态置为"上报失败：5"并记下上报失败原因到返回信息中，上报完成后主表数据状态置为"已完成：3"，以上操作都在一个事务里。 （对于不需人工审核的报告，上报成功即为上报完毕，对于需要人工审核的报告上报后仍需等待审核人员通后才算上报完毕，目前该数据平台系统暂不支持审核反馈和相关查询。）

## 应答文件

对各子表进行遍历，组装整合各报告的返回信息和接收人编号，生成一般应答文件并发送到指定位置（单文件夹根据名字区分），发送完成后，将相应主表状态置为"已回执：4"。

如需增值服务还可生成详细错误应答文件发送到指定位置。

应答文件详情见1.3.2通用应答文件示例。

# 渠道控制

目前数据平台只能单一控制用户是否能够通过文件进行上报，对于报告库页面上报无法控制，要想控制所有上报渠道需要报告库前台与后台的共同改造才能实现。报告库前台增加上报权限控制和数据平台访问接口，后台增加渠道来源。通过报告库的权限控制和数据平台的交互来完成渠道报送控制。

# 异常机制和重报机制

在各流程过程中如果出现异常情况会记录到日志并发送警告到运维和开发人员手中。如上报过程中等待超时或通信中断会触发异常，记录信息为异常发送给相关人员。

报告可以进行手动重报，通过页面或数据库状态操作。

# 5.数据统计与展示

数据入库存储后可根据状态时间等信息统计上报情况，开放相应查询接口以供其他系统调用查询。

校验库管理系统也可开放接口给报告库系统调用。