# 数据平台方案

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\_131101A01\_000899\_A0001\_20171211\_0001\_A.TXT

OTC\_131101A01\_000899\_A0001\_20171211\_0002\_A.TXT

* + 1. **库表设计**

1. **文件内容表** **handle\_plain\_tmp**

存储解压后文本文件内容条目信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| file\_id | int | 20 | 处理文件表主键 |
| submission\_record\_id | varchar | 50 | 报送记录号 |
| file\_item\_content | longtext | 0 | 文件条目内容 |
| file\_item\_sort | int | 20 | 文件条目序号 |
| data\_json | longtext | 0 | json数据 |
| create\_date | datetime | 0 | 创建时间 |
| update\_date | datetime | 0 | 修改时间 |

1. **文件内容元组表** **handle\_tuple\_tmp**

存储解压后文本文件内容元组信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| file\_item\_id | int | 20 | 文件内容表主键 |
| tuple\_name | varchar | 100 | 元组名称 |
| submission\_record\_id | varchar | 50 | 报送记录号 |
| tuple\_item\_content | longtext | 0 | 元组条目内容 |
| tuple\_item\_sort | int | 20 | 元组条目序号 |
| create\_date | datetime | 0 | 创建时间 |
| update\_date | datetime | 0 | 修改时间 |

1. **字段索引表** **handle\_field\_index\_tab**

存储接口字段与文件内容中索引值对应信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| t\_template\_type | varchar | 255 | 模板编号 |
| t\_element\_id | varchar | 255 | 元素名称 |
| t\_element\_name | varchar | 255 | 业务字段名称 |
| d\_field | varchar | 255 | 数据库字段名称 |
| d\_tuple\_up\_name | varchar | 255 | 上级tuple数据库字段名称 |
| t\_control\_type | varchar | 255 | 控件类型  0|普通元素、1|Tuple子项元素、  2|多附件、3|单附件、  4|Tuple父元素（虚元素） |
| index\_value | int | 10 | 字段存储的索引值 |

1. **处理压缩包表** **handle\_compress\_package\_tab**

存储压缩文件等信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| zip\_file\_name | varchar | 100 | 压缩包文件名 |
| status | int | 2 | 状态 |
| create\_date | datetime | 0 | 创建时间 |
| update\_date | datetime | 0 | 修改时间 |

1. **处理文件表** **handle\_file\_tab**

存储文本文件等信息

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| zip\_id | int | 20 | 压缩包主键 |
| zip\_file\_name | varchar | 100 | 压缩包文件名 |
| 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 |
| local\_error | varchar | 2000 | 本地校验错误 |
| 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

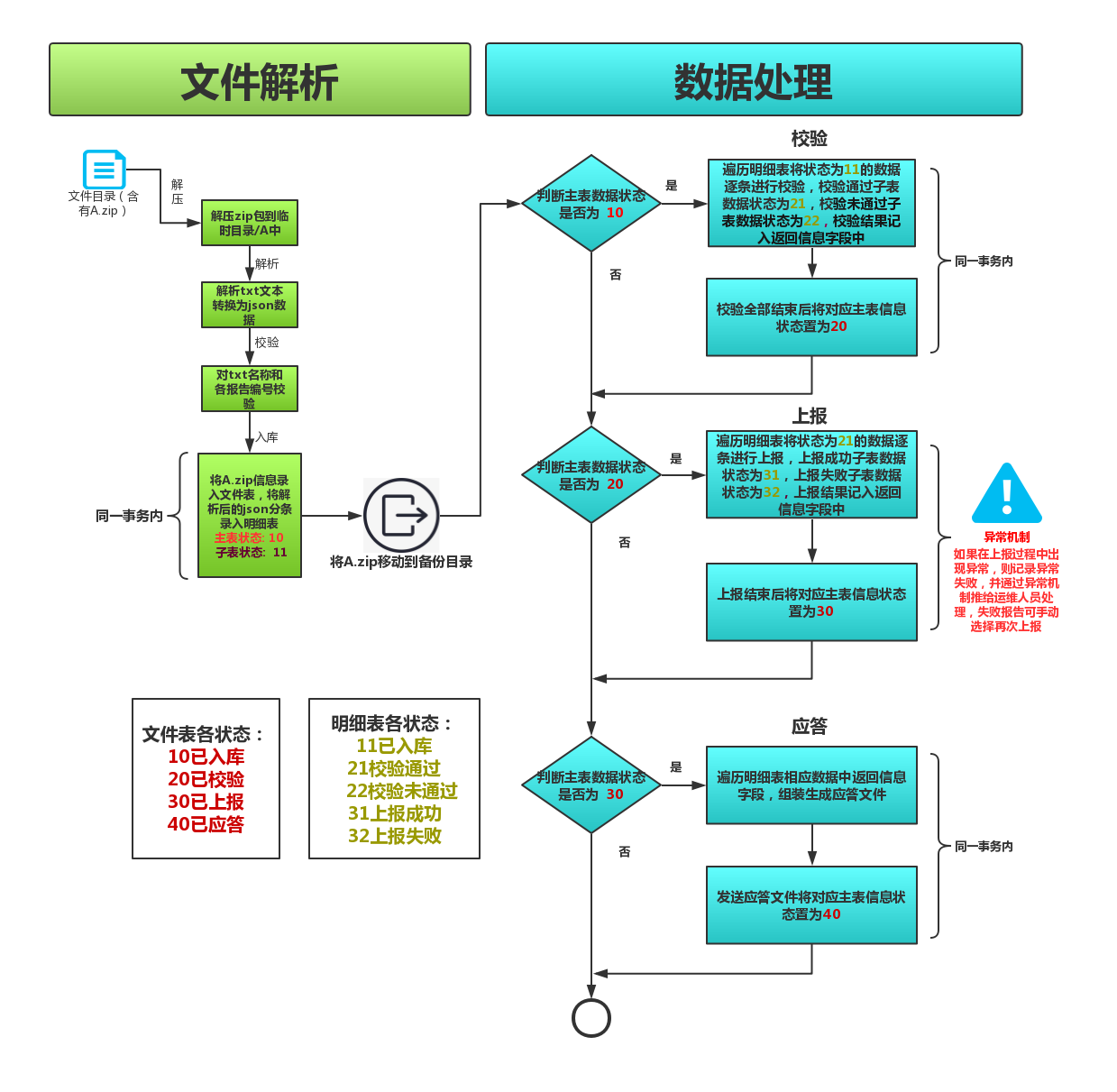
|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 类型 | 长度 | 注释 |
| id | int | 20 | 主键 |
| file\_id | int | 20 | 处理文件表id |
| file\_creater\_code | varchar | 10 | 文件创建人代码 |
| file\_receiver\_code | varchar | 10 | 文件接收人代码 |
| interface\_id | varchar | 10 | 接口标识 |
| submission\_date | varchar | 10 | 报送日期 |
| file\_num | varchar | 10 | 文件序号 |
| operation\_id | varchar | 2 | 操作标识 |
| submission\_record\_year | varchar | 50 | 年份 |
| submission\_record\_month | varchar | 50 | 月份 |
| submission\_record\_id | varchar | 50 | 报送记录号 |
| data\_json | longtext | 0 | json数据 |
| status | int | 2 | 状态 |
| submission\_msg | varchar | 2000 | 报送返回提示信息 |
| submission\_error | varchar | 2000 | 报送返回错误信息 |
| submission\_bizid | varchar | 2000 | 报送返回报告业务编号 |
| submission\_protcolid | varchar | 2000 | 报送返回协议编号 |
| submission\_data | varchar | 2000 | 报送返回附加信息 |
| response\_msg | varchar | 2000 | 校验信息 |
| local\_error | varchar | 2000 | 本地校验错误 |
| 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 |
| zip\_id | int | 20 | 处理压缩包表id |
| file\_id | int | 20 | 处理文件表id |
| logging\_type | int | 2 | 日志类型  10:本地  20:校验库  30: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**
* **解析入库**

扫描目录文件（zip）代号为A.ZIP，获取A.ZIP名称信息的业务用户（业务用户9位），获取A.ZIP名称信息，数据**状态**为"已入库：10"，保存录入**处理压缩包表**，解压A.zip文件到临时目录/A中，遍历临时目录/A下的文件列表，将文件名取出匹配文本名正则表达式，若匹配则获取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)，若**文件接收人代码**不是000899则数据**状态**为"本地校验失败：60"，**本地校验错误**为(1013："文件接收人代码与平台代码不符")，若**操作标识**不是A则数据**状态**为"本地校验失败：60"，**本地校验错误**为(1014："操作标识错误，应为A")；保存录入**处理文件表**，比对B（文件创建人代码9位）与A.ZIP名称信息的业务用户（业务用户9位），若不一致则数据**状态**为"本地校验失败：60"，**本地校验错误**为(1001："文件名中的机构号与压缩文件中的不一致")；若一致则获取B.txt文件里的接口标识判断是否在接口范围之内，若不在范围内则数据**状态**为"本地校验失败：60"，**本地校验错误**为(1003："接口标识不在接口范围之内")；若在接口范围之内则获取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，则数据**状态**为"本地校验失败：60"，**本地校验错误**为(1000："该报告文件已上报过，请勿重复上报")；若不存在，则校验B.txt文件，若校验失败则数据**状态**为"本地校验失败：60"，**本地校验错误**为(1008："文本内容不合规")；若校验成功则数据状态为"已入库：10"，解析B.txt文本录入到**文件内容表与文件内容元组表**；若文件列表为空则记录**系统日志表(压缩包id)**(1009："该压缩包下没有文件")；若文件列表一个也没有匹配到文本名称正则表达式则记录**系统日志表(压缩包id)**(1010："该压缩包下没有匹配的文件，请检查文本文件名称")；遍历临时目录/A下的文件列表完毕后，将临时目录/A复制到附件目录/A下，将A.zip移动到备份目录中，删除临时目录/A；以上操作没有事务。

* **比对入库**

查询**处理压缩包表**数据**状态**为"已入库：10"的数据，遍历**处理压缩包表**数据，根据**压缩包主键**查询**处理文件表**数据**状态**为"已入库：10"的数据，遍历**处理文件表**数据，根据**主键**查询**文件内容表与文件内容元组表**，若**文件内容表**数据不为空则遍历**文件内容表**数据，取出**报送记录号**和**处理文件表主键，**查询数据是否大于1，大于1则录入**处理明细表状态**变为"本地校验失败：12"，**本地校验错误**为(1016："该报告记录号在文件中重复，请核对")；否则取出**报送记录号**和**处理文件表(文件创建人代码、文件接收人代码、接口标识)**、**状态**为"上报成功：31"查询**处理明细表**数据，若不存在则为数据正常入库；若存在则为数据比对入库；**文件内容表**数据循环完毕后，**处理文件表**数据**状态**置为"已比对：50"，删除**文件内容表与文件内容元组表**数据；若**文件内容表**数据为空则**处理文件表**数据**状态**为"本地校验失败：60"，**本地校验错误**为(1007："该报告文件文件内容未识别，请仔细核对")；**处理文件表**数据循环完毕后，**处理压缩包表**数据**状态**置为"已比对：50"；查询**处理文件表**数据为空，**处理压缩包表**数据**状态**置为"已上报：30"，以上操作都在一个事务里

**数据正常入库**

**文件内容表**数据录入**处理明细表**，数据**状态**为"已入库：11"，若数据中元组不存在，则**状态**变为"本地校验失败：12"，**本地校验错误**为(1015："该报告下不存在元组")；若数据转化为报送数据(json)时异常，则**状态**变为"本地校验失败：12"，**本地校验错误**为(1004："数据转化为报送数据异常")；若报送记录号为空，则**状态**变为"本地校验失败：12"，**本地校验错误**为(1005："报送记录号为空异常")。

**数据比对入库**

**处理文件表**数据移植进**处理明细表**，数据**状态**为"本地校验失败：12"，**本地校验错误**为(1006："该报告记录号已上报成功，请勿重复上报")。

* **校验**

查询**处理压缩包表**数据**状态**为"已比对：50"的数据，遍历**处理压缩包表**数据，根据**压缩包主键**查询**处理文件表**数据**状态**为"已比对：50"的数据，遍历**处理文件表数据，**根据**主键**查询**处理明细表**数据**状态**为"已入库：11"进行校验，遍历**处理明细表**数据，获取文件创建人代码（截取前6位）与报送记录号（截取前6位）进行比对，若比对不一致则**处理明细表**数据**状态**置为"本地校验失败：12"，**本地校验错误**为(1002："该报告报送记录号中机构号与文件名中的不一致")，若比对一致则开始调用校验接口，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**校验信息(**未通过原因**)**，**处理明细表**数据遍历完毕后，**处理文件表**数据状态置为"已校验：20"，**处理文件表**数据循环完毕后，**处理压缩包表**数据**状态**置为"已校验：20"；查询**处理文件表**数据为空，**处理压缩包表**数据**状态**置为"已上报：30"，以上操作都在一个事务里，若调用校验接口出现异常情况，则记录**系统日志表(压缩包id、处理文件id)**。

* **上报**

查询**处理压缩包表**数据**状态**为"已校验：20"的数据，遍历**处理压缩包表**数据，根据**压缩包主键**查询**处理文件表**数据**状态**为"已校验：20"的数据，遍历**处理文件表数据，**根据**主键**查询**处理明细表**数据**状态**为"校验通过：21"开始调用上报接口进行上报，遍历**处理明细表**数据，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**报送返回错误信息**(上报失败原因)，**处理明细表**数据遍历完毕后，**处理文件表**数据**状态**置为"已上报：30"，**处理文件表**数据循环完毕后，**处理压缩包表**数据**状态**置为"已上报：30"；查询**处理文件表**数据为空，**处理压缩包表**数据**状态**置为"已上报：30"，以上操作没有事务，若调用上报接口出现异常情况，则记录**系统日志表(压缩包id、处理文件id)**。

* **应答**

查询**处理压缩包表**数据**状态**为"已上报：30"的数据，遍历**处理压缩包表**数据，获取**压缩包名**,生成应答目录/A，根据**压缩包主键**查询**系统日志表**，若存在，则获取日志信息生成应答文件至应答目录/A中；若不存在，则根据**压缩包主键**查询**处理文件表**，遍历**处理文件表数据**，根据**处理文件主键**和**压缩包主键**查询**系统日志表**，若存在，则获取日志信息生成应答文件至应答目录/A中；若不存在，则判断数据**状态**若为"本地校验失败：60"，则获取本地校验错误生成应答文件至应答目录/A中；若数据**状态**为"已上报：30"，则根据**主键**查询**处理明细表**数据，遍历**处理明细表**数据生成应答文件至应答目录/A中，**处理明细表**数据遍历完毕后，**处理文件表**数据**状态**置为"已应答：40"，**处理文件表**数据循环完毕后，**处理压缩包表**数据**状态**置为"已应答：40"；删除附件目录/A，打包应答目录/A，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

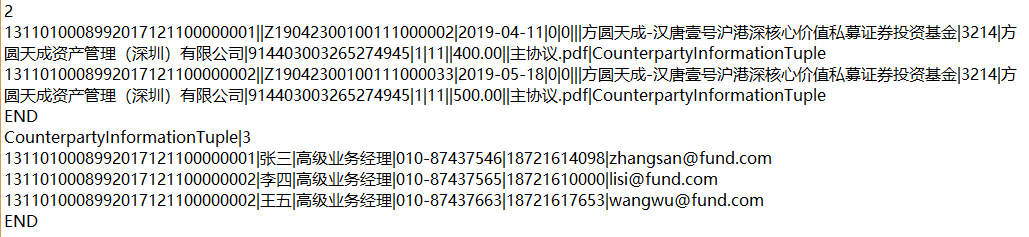
接收方编号 000899

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。

* + 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.扩展与补充

数据入库存储后可根据状态时间等统计上报情况，如有需要也可通过页面进行展示。