



集算器

创新大数据计算引擎

某税务局数据手册流程自动化

润乾软件出品

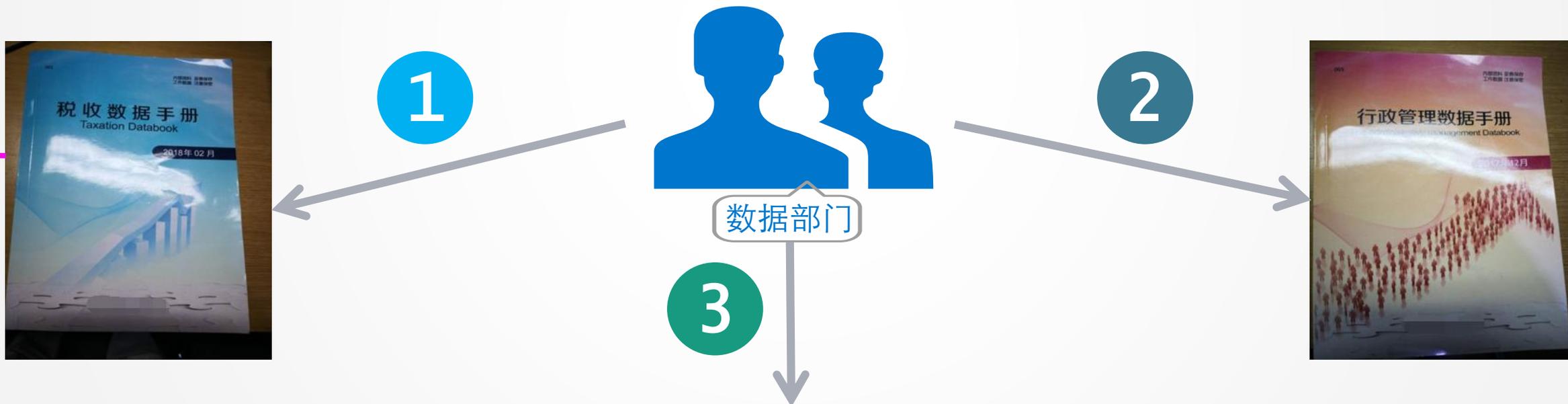




项目背景

“税收数据手册”主要工作是税务数据部门定期汇集各业务部门的发布表样及业务口径，要求运维公司等提供表中数据内容，由业务部门审核确认，并提交给印刷厂排版印制，最后获得纸质文档，通过分发、邮寄、送交等手段按期交给领导、上级、兄弟部门。每月一期，在次月15日前需印制完毕。

数据部门还承担“行政管理数据手册”的工作。行政管理数据手册主要工作涉及税务部门人、财、物的管理，每半年生成一次。



为提高部门人员的工作效率，简化、减轻工作流程，客户引入了报表二次开发项目，实现了手册数据生成、审批、数据审核、excel导出等电子化目标，并于2015年正式上线！



原方案存在的不足



业务不稳定

- 由于数据手册涉及的业务部门众多，税收政策多变，领导关注的管理点也多变，这就造成手册业务的极度不稳定，几乎每期都会发生需求变更

服务常态化

- 为应对业务不稳定性，客户需不断购买运维服务，以确保在出表期有技术人员常态配合，随时根据业务需求变动而修改应用程序，确保印刷出的手册成品的质量和内容，符合业务部门的标准

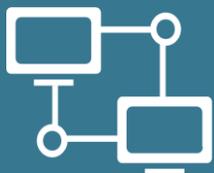
客户投诉多

- 被客户投诉内容包含表样格式不正确、数据不满足审核关系等。导致已付印的书籍报废，客户方工作显得很被动。领导甚至怀疑有人员不认真、故意出错，项目组成员承受较大压力

洞察现象后的本质



为彻底解决上述问题，公司高层决定投入人力，在不改变原有流程的情况下，对出表后的审核流程、确定提交印厂的流程进行优化！原因总结如下：



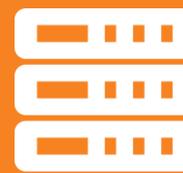
临时性需求占比很大，而报表开发过程中更多的困难在数据源准备上，所以来不及开发部署上线



数据源与报表耦合度高，一次临时的表样变动，就会造成数据源、应用部署等修改范围联动，维护困难



手工操作过多，无论线上出表、多人审核、数据调整、传输等都是手工操作，环节琐碎、出错概率大



印制前虽会手工核对，但对于格式出错、数据误贴、标题注释漏改等情况，完全凭借人员经验及责任心

新目标的提出



在不改变现有流程的情况下，建立“数据手册桌面版”工具，覆盖从excel文档提数到提交印厂的整个过程！期望能够建立长期、低成本、快速的应对机制！

桌面级应用

不依赖于外部网络、数据库等环境。基于文件系统的目录构建，可按月保存，基于上次的版本快速修改调整。

过程可追溯

数据提取、计算数据加工、审核数据、输出数据等，通过目录保存，确保整个过程可追溯。

快速稳定高效

执行程序要在10分钟之内跑完全过程，操作员需提前设定好源数据(excel文件)及各种审核、计算规则。

excel表外数据处理

把数据(表中的数据集、标题、注释等)从excel中分离出来，以此为基础进行数据计算、数据审核、文档输出。

一键自动化

直接由程序生成excel套表文件及adobe indesign文档。

降低技术门槛

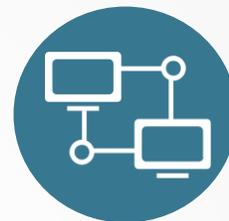
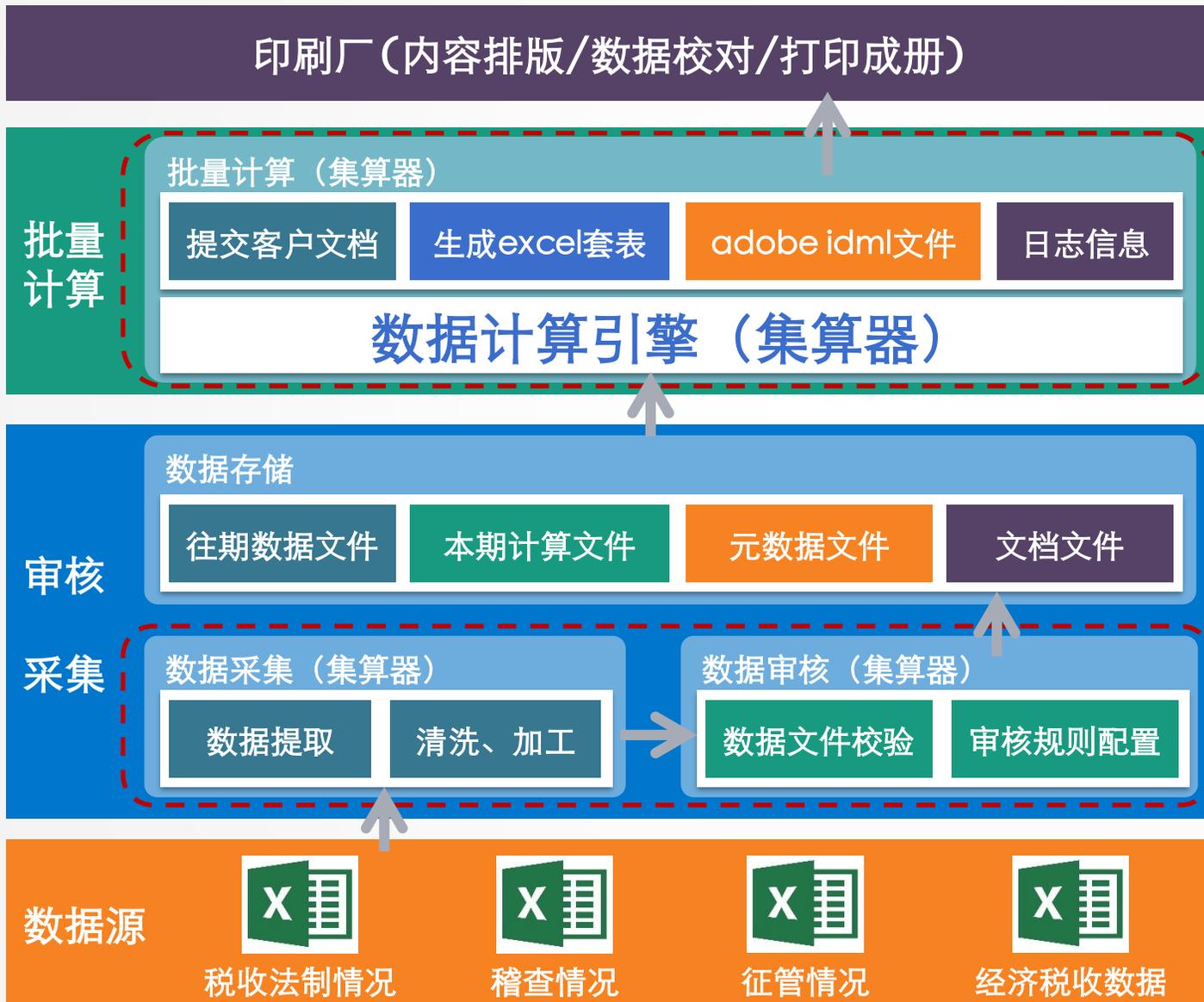
新增、修改各种表格、口径规则调整尽量通过配置文件、元数据文件等实现，不需要运维人员有强开发能力。





引入数据计算引擎中间件

税收数据手册桌面版工具



大幅缩短工期

丰富高效类库、敏捷语法体系，完备的开发、调试功能，让excel计算更有效率



— “人” 饰多角

数据采集、数据审核、批量生成等，其本质都是数据计算，可由集算器统一完成



绿色桌面应用

无框架体系，标准接口便于实现应用嵌入式集成，应用部署更容易

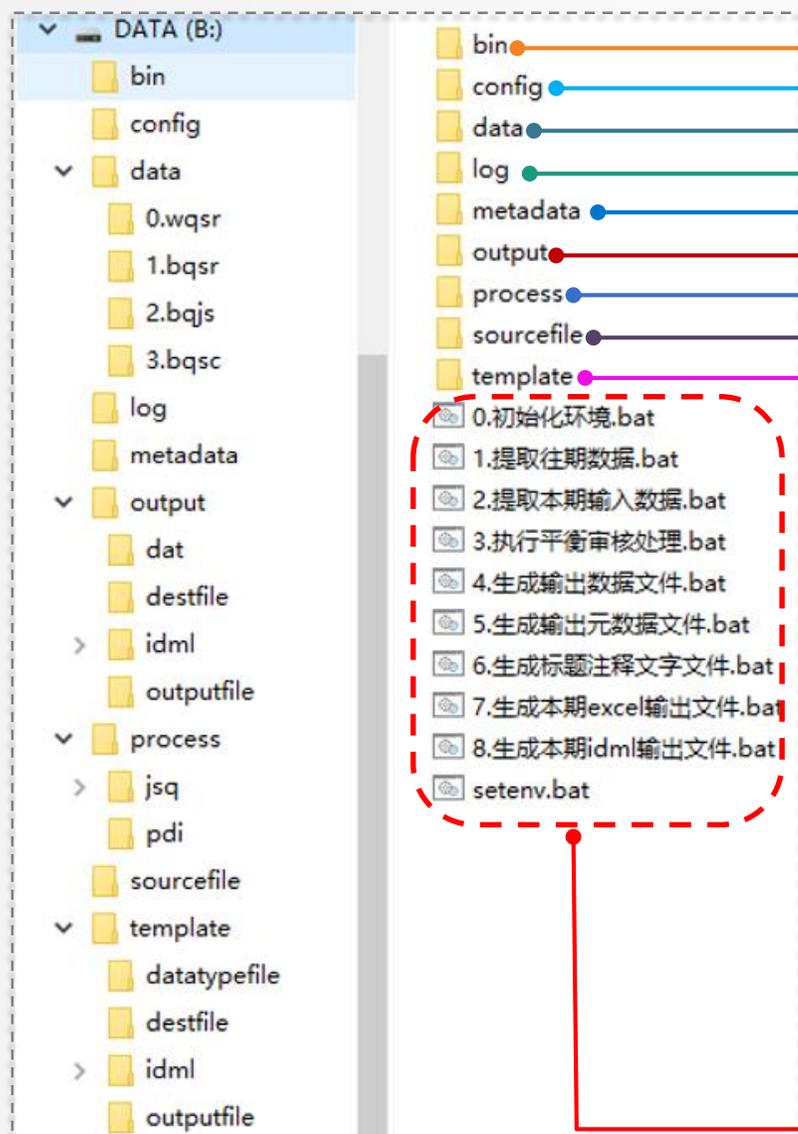


真正随需而动

脚本文件化易于管理，需求更变无须编译，一般技术人员都能做，成本低，响应快



应用程序简述



1 bin放置需要的执行程序，当前包含zip.exe。

2 config目录放置scjbx.ini文件，其中包含数据文件目录、处理年月等信息。

3 data目录包含往期输入、本期输入、本期计算、本期输出等，存放各环节生成的数据文件。

4 log目录存放运行日志。

5 metadata目录即元数据目录，通过xlsx文件存放各种规则数据。是运维人员主要调整的目录。

6 output目录即输出目录，存放生成需提交客户、印厂的excel套表、adobe idml文件等。

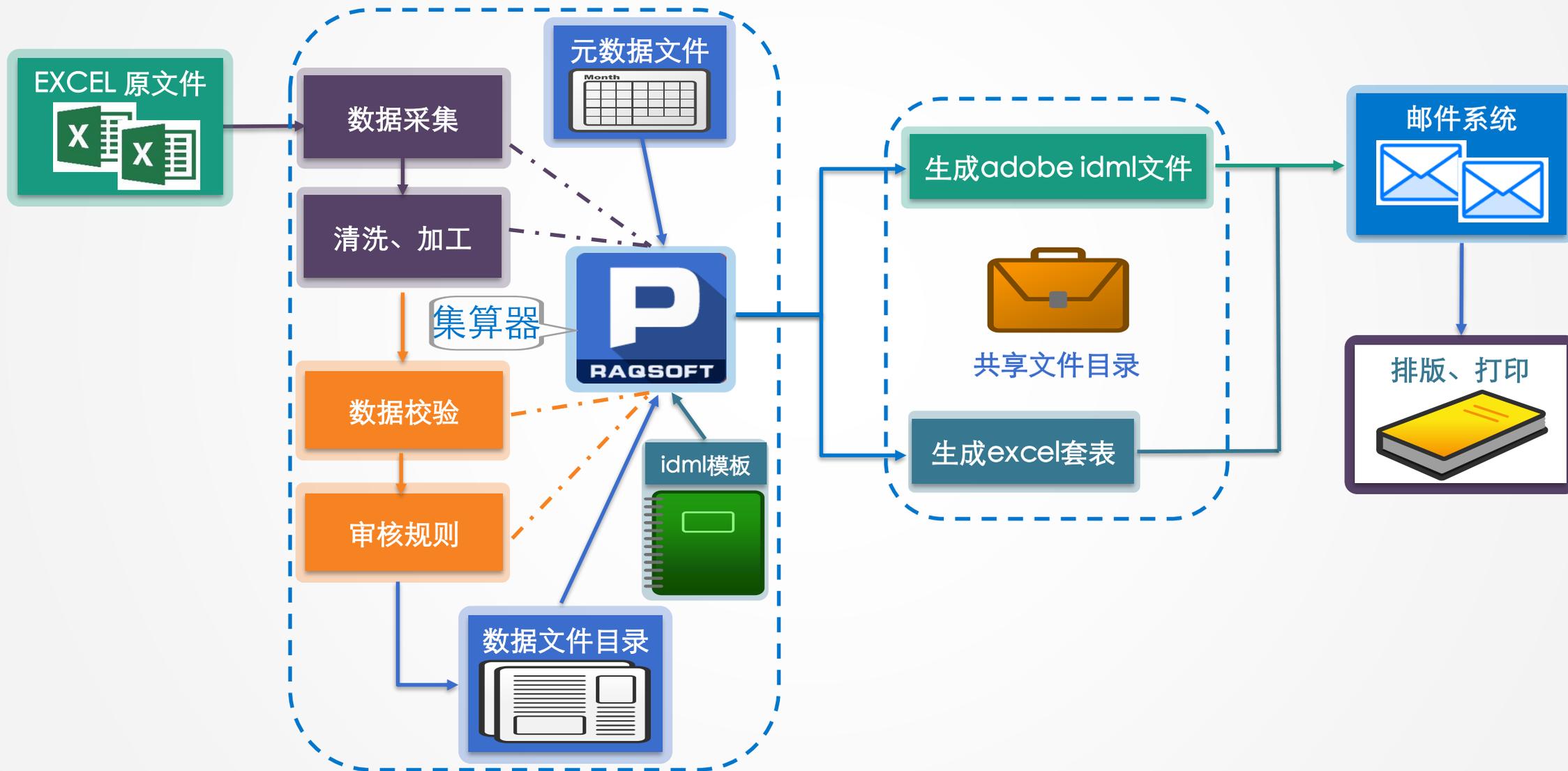
7 process目录存放所有数据计算的算法逻辑文件(集算器)。

8 sourcefile目录，存放运维人员获得的各种excel源数据文件，以及描述表样与数据文件。

9 template目录，存放与输出相关的excel模板及idml模板目录。

0 整个环境建立、数据提取、加工、审核、输出、执行数据文件、文档文件、日志文件等按照上述要求放在相关目录中。

数据计算流程



示例：集算器代码与输出文件



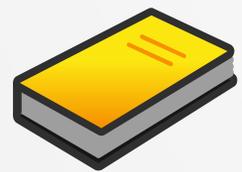
一段审核程序的举例



```

1  ="B:/config/scjbx.ini"      =file(A1)      =file("B:/log/sh_log_"+string(now(),"yyyyMMdd_HHmm")+".xlsx")
2  ="B:/metadata/sh_list.xlsx"  "shid"        9:             id,shid,qybs
3  "datfile"                  9:             shid,filename,ds,dir,qybs
4  "dfx"                       9:             id,dfx,qybs
5  ="B:/metadata/zb_list.xlsx"  "Sheet1"      9:             id,mc
6  =file($A$2)                 =create(序号,审核id,表名,类别,文字) >$C$1.exportxls@t($B$6)
7  =A6.importxls@t($D$2);$B$2  =A6.importxls@t($D$3);$B$3  =file($A$5).importxls@t($D$5);$B$5,$C$5).keys(i
8  =create(ds,nr)
9  =output@t("执行通用审核程序开始")
10 for D7
11   =$A$10.mc
12   =output@t($B$10,$C$10,"开始") =true
13   for $D$10
14     =$B$1.property($B$12.dir+$B$12.filename+".dat") =file($C$12)
15     if $D$12.exists()
16       >A8.insert(0,$B$12.ds,$D$12.import@timport@t(",|"))
17     else
18       =output("执行错误",$C$10,$C$12,"不存在")
19       >$C$11=false
20       break
21     if $C$19.count()>0 >$C$1.exportxls($C$19)
22   >$A$8.reset()
23   =output@t($B$10,$C$10,"结束")
24 =output@t("执行通用审核程序结束")
25 =$B$1.property("para_process_sh")
26 =A6.importxls@t($D$4);$B$4,$C$4).select(qybs=="Y")
27 =output@t("执行dfx审核程序开始")
28 for A25
29   =$A$24+$A$27.dfx+".dfx"
30   =output@t($A$27.id,$B$27)
   =call($B$27)
   =output@t("执行dfx审核程序结束")
    
```

30行



输出idm1文件并打印成册

序号	项目	总户数	正常户	非正常户	停业户	在途户	有税户		无税户	机构审批部
							户数	占正常户比重%		
1	合计	1799248	1778296	20952			1062650	59.8	715646	
2	一、法人	1290343	1273627	16716			964967	75.8	308660	
3	1.企业法人	1270945	1254305	16640			949265	75.7	305040	工商局
4	2.事业单位法人	13701	13677	24			11294	82.6	2383	编办
5	3.社会团体法人	3868	3820	48			2663	69.7	1157	民政
6	4.机关法人	1829	1825	4			1745	95.6	80	编办
7	二、其他	508905	504669	4236			97683	19.4	406986	
8	1.分支机构	55969	55598	371			43116	77.5	12482	工商局
9	2.民办非企业单位	2522	2505	17			2154	86.0	351	民政
10	3.个体工商户	446585	442746	3839			49414	11.2	393332	工商局
11	4.律师事务所	2500	2496	4			2238	89.7	258	司法局
12	5.其他组织	1329	1324	5			761	57.5	563	农经办

注释：1. “有税户”为2017年缴纳过税款的户数
2. 由于系统中登记环节无数据标识，数据查询量较大，故目前报表更新频率为按半年更新，待系统完善后，再按月更新。

用户评价

集算器利用较少的代码（一般低于30行），能够实现极具表现力的效果；特有的宏机制极大地提高了代码复用程度，进一步减少工作量。可以说不用集算器，很难用其他程序在同等的时间开发出当前功能的应用！

应用效果

- 1 实现了一个通用的数据提取、计算、审核校验、输出框架。以此为基础，其实可以覆盖这样的一类需求，具有一定的普遍性意义。
- 2 极大程度降低对项目人员技能要求。运维人员只要大致明白原理，然后专注在业务变化本身，功能实现使用excel设置即可。
- 3 通过配置文件、元数据形成可扩展性。技术上该程序只依赖java环境，对操作系统、office程序、adobe应用都没有依赖。事实上程序换到linux环境也无需做多大调整，熟悉的开发人员一个工作日即可实现。

集算器凭什么？



独立计算引擎

不依赖于数据库的计算能力，拥有完整的敏捷语法体系



敏捷语法体系

语法简洁易学，更接近人的自然思维，特别适合做复杂过程计算



解释执行语言

应对没完没了的临时性需求，即开即用，无须编译或重启应用



多样性数据源

丰富的多样性数据源接口，直接取数计算

创新技术

推动应用进步!

