
传输交换管理平台
技术白皮书

亿信华展

ESENSOFT

目录

1	前言	4
1.1	关于本白皮书.....	4
1.2	背景介绍.....	4
2	产品概述	4
3	产品架构	5
4	产品优势	6
4.1	通用性.....	6
4.2	可扩展性.....	6
4.3	易用性.....	6
4.4	安全性.....	7
4.5	高性能.....	7
4.6	可靠性.....	7
5	产品技术特点.....	7
5.1	产品总体特性.....	7
5.1.1	产品架构成熟通用.....	7
5.1.2	开放的二次开发接口.....	7
5.1.3	可扩展和集成性.....	8
5.1.4	资源库隔离机制.....	8
5.2	数据建模.....	8
5.2.1	支持可视化建模.....	8
5.2.2	建模方式丰富.....	8
5.2.3	模型属性扩展管理.....	8
5.3	数据传输交换.....	9
5.3.1	可视化数据交换作业设计器.....	9
5.3.2	数据源支持.....	9
5.3.3	跨数据源取数.....	9
5.3.4	数据交换过程调试方式.....	9
5.3.5	增量捕获方式.....	9
5.3.6	数据分发方式.....	9
5.3.7	丰富的交换组件.....	10
5.3.8	清洗转换规则.....	10
5.3.9	数据校验.....	10
5.4	调度管理.....	10
5.4.1	交换任务流程设计器.....	10
5.4.2	交换任务组件支持.....	11
5.4.3	交换任务流程发布.....	11
5.4.4	任务调度服务.....	11
5.4.5	交换任务导入/导出.....	11
5.5	系统监控.....	11
5.5.1	任务运行监控.....	11
5.5.2	系统运行监控.....	12

5.6	系统管理.....	12
5.6.1	备份与恢复.....	12
5.6.2	系统审计日志.....	12
6	产品功能.....	12
6.1	参数配置.....	12
6.1.1	采集参数.....	12
6.1.2	分发参数.....	13
6.1.3	监控参数.....	13
6.1.4	全局参数.....	13
6.2	数据处理.....	14
6.2.1	作业设计器.....	14
6.2.2	输入输出组件.....	14
6.2.3	转换组件.....	15
6.2.4	脚本组件.....	16
6.2.5	增量识别.....	16
6.2.6	文件拆分.....	18
6.2.7	文件合并.....	19
6.2.8	一致性维度.....	21
6.2.9	数据清洗.....	22
6.3	数据调度.....	23
6.3.1	数据传输.....	23
6.3.2	数据压缩.....	25
6.3.3	断点续传.....	25
6.3.4	数据加密.....	26
6.3.5	一致性验证.....	26
6.3.6	异常报警.....	27
6.3.7	调度配置.....	27
6.3.8	调度监控.....	28
6.4	运营监控.....	29
6.4.1	系统监控.....	29
6.4.2	作业监控.....	30
6.4.3	监控配置.....	30
6.5	用户及权限管理.....	30
7	应用案例.....	错误!未定义书签。
7.1	税务行业.....	错误!未定义书签。
7.2	卫生行业.....	错误!未定义书签。
7.3	法院行业.....	错误!未定义书签。
8	运行环境.....	错误!未定义书签。
8.1	硬件环境保障.....	错误!未定义书签。

1 前言

1.1 关于本白皮书

本白皮书对应产品版本为：传输交换管理平台V2.1。

最后修订日期：2019年07月。

本白皮书将在阐述数据传输交换管理平台的背景和意义的基础上，详细介绍数据传输交换管理平台在技术和功能上的特点。

1.2 背景介绍

随着企业综合管理水平的提高和信息化建设的开展和深入，企业间各类业务数据之间的交换成为了越来越突出的问题。从全局范围来看，越来越多的企业采用数据总线架构统一数据交换通道，以避免各应用之间繁冗杂乱的数据接口。但是，从目前的信息化现状来看，企业内部的计算机应用系统通常是各个部门的情况和需求独立建设的，所采用的软件技术、数据格式等存在较大差异，缺乏统一的规划和标准，各系统之间的信息共享和交互较为困难，传输协议和数据格式五花八门，企业应用网内外隔离，涉密数据的安全因素等问题都是我们所面临的难题。

传输交换管理平台应需而生，通过传输交换管理平台对企业内部应用系统进行统一规划，利用底层整合的信息资源，以实现各部门业务数据在应用层面的互联互通和信息共享，提供增量识别、数据清洗、数据传输、数据压缩、一致性验证等多种手段保证数据的接入，规范数据流转，提升数据共享水平，更加有效发挥数据资产价值，并提供多套数据脱敏方案和加密算法保证数据的安全性。

2 产品概述

传输交换管理平台是华宇信息独立研发的具有自主知识产权、业界领先的企业级数据整合应用平台。通过多年的项目实践，积累了大量的行业解决方案。目

前被广泛应用于金融、政府、卫生、法院等行业的数据中心项目。

传输交换管理平台 2.1，为传输交换管理平台 2.0 的升级版本，优化了传输性能、增加了部分功能、提高了安全性。

3 产品架构

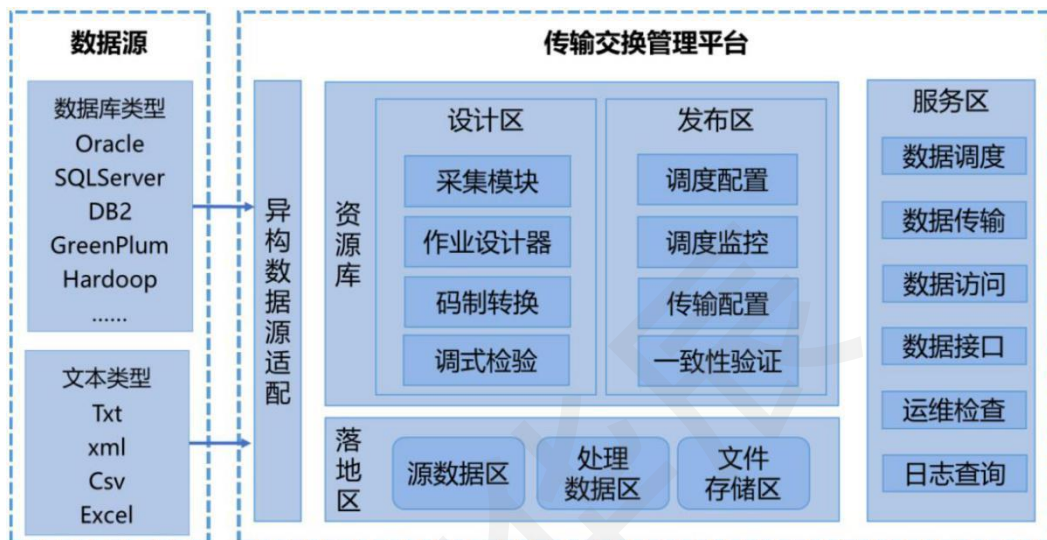


图1 产品架构图

数据源层，支持主流的数据库源、平面文本文件、Excel 文件和其它格式数据。

异构数据源适配层，基于数据库和服务器的内存方式，完成异构数据源的适配、整合和交叉访问。

落地地区，包括源数据区、处理数据区、及文件存储区。

资源库，分为运行区和设计区，运行区是数据运行环境，提供传输、调度控制服务，作业实时运行监控和系统服务质量监控，设计区是数据交换作业流程设计环境，内置码制转换、调式校验等控件

服务区，开放数据访问服务，提供 webservices、api 等多种接口服务，直接面向运维和浏览用户进行查询。

4 产品优势

4.1 通用性

该产品采用面向对象语言设计，安装部署与硬件、操作系统无关，广泛的数据采集类型（数据库、文件、表格等）兼容，适用于数据库之间、文本之间、以及不同类型之间进行数据交换等数据交换场景，多种数据变化捕获方法完成数据的增量抽取，丰富的交换组件和清洗转换规则支持，全方面满足数据加工、数据交换和数据传输需求。

4.2 可扩展性

产品内置了多种函数库，提供跨数据库类型和版本的算法兼容性，同时支持自定义，提供了丰富的脚本组件、转换组件、清洗组件、存储过程组件、系统命令组件的方式来保证工具的扩展性。

4.3 易用性

采用JAVA、J2EE模式构建，基于Web瘦客户端架构。服务器应用部署完毕，在客户端无须安装任何插件，就能直接访问。

同时在设计上充分结合大量项目实施习惯和场景要求，来保证工具的便捷性和易用性。主要包括：

- 敏捷开发模式，集数据建模和 ETL 设计于一体，能快速响应开发，灵活应对需求变动。
- 多种转换组件满足所有数据交换中对数据的转码、清洗、拆分、合并等处理。
- 开发者与应用者视角隔离，支持多人协同定义，共享资料库。
- 可视化模型定义、交换任务过程定义、调度过程定义。
- 支持数据抽样测试，数据交换过程支持边设置边预览。

4.4 安全性

采用加密传输机制，对数据的传输进行加密，内置大量的数据脱敏算法，支持对系统和数据的操作行为审计，来保证数据的安全性。

4.5 高性能

支持数据分区、作业并行和数据并行装载技术来保证数据采集、传输和分发的高效性。支持大量的基于数据库或操作系统脚本语言和技术，来提供数据库端的数据采集和分发性能。

4.6 可靠性

支持数据缓存、断点续传和错误通知、调度监控等技术，保证工具使用的稳定性和可靠性。

5 产品技术特点

5.1 产品总体特性

5.1.1 产品架构成熟通用

总行服务端采用 J2EE 架构和 B/S 模式，采用 JAVA 语言开发，支持 Tomcat、WebLogic 等多种应用服务器运行平台。用户可以通过 IE，chrome 等浏览器进行访问。

5.1.2 开放的二次开发接口

传输交换管理平台是基于我公司统一的公共研发平台进行搭建的，提供二次开发接口，可以根据需求方便地对产品进行灵活定制修改和功能扩展。

5.1.3 可扩展和集成性

具备良好的可扩展性和集成性，提供丰富的资源访问接口技术，支持如 API 接口、URL 接口、Web Service 接口等第三方接口调用。

5.1.4 资源库隔离机制

产品内置两个资源库，分为设计区和运行区，两个资源库相互隔离，可保证运行环境的稳定和可靠，所有作业的修订和更改通过在设计区完成调试和试运行后，通过发布机制，发布到稳定的运行区，在生产环境中可区分模拟运行和正式上线环境，建设系统运行风险。

5.2 数据建模

5.2.1 支持可视化建模

软件支持图形方式完成数据模型创建，可以在可视化建模界面中新建、编辑、删除数据模型，可视化建模窗口能直观的展现模型与模型之间的关系。

5.2.2 建模方式丰富

软件除了支持常规的手工自定义创建模型之外，还应支持拾取数据库表、根据已有的数据库表字段、数据库数据集等多种方式创建数据模型。

5.2.3 模型属性扩展管理

软件在管理和维护数据模型属性时，除了支持常规的物理属性的管理也支持逻辑业务属性的管理，并提供统一的属性维护界面，来维护属性的展现方式和是否支持检索等信息，可由用户自定义扩充。

5.3 数据传输交换

5.3.1 可视化数据交换作业设计器

支持拖拽的方式完成整个交换作业流程设计，在设计器面板中支持拖拽数据模型、交换组件、连线来完成交换过程的定义和维护，设计器支持放大缩小、保存、自动布局、参数设置、语法检查、撤销等基本功能。

5.3.2 数据源支持

产品提供丰富的数据源接口，除支持主流的数据库外，还支持文本文件、SQL 结果集、接口数据源等多种形态的数据源。

5.3.3 跨数据源取数

支持从异构数据源取数，能直接完成跨数据库的数据关联取数。

5.3.4 数据交换过程调试方式

软件支持对数据交换过程进行运行到、顺序执行、模拟运行等多种方式的调试，提供详细的日志信息来记录交换任务每个步骤的运行时长和状态信息，在图形窗口提供状态提醒和步骤出错说明。

5.3.5 增量捕获方式

产品支持时间戳、MD5、触发器、全表比较等多种方式的增量数据提取方式，来捕获数据变动信息，完成增量数据的落地。

5.3.6 数据分发方式

数据分发提供多种方式来满足数据应用的要求，如：数据追加、数据覆盖、

数据更新、更新插入等。数据追加用于往目标数据末尾写入新数据，数据覆盖清空数据完成数据的全覆盖、数据更新会更新目标中存在的数，更新插入会写入新增数据更新存在数据。

5.3.7 丰富的交换组件

支持丰富的输入输出组件、转换组件、脚本组件等，输入输出组件用于完成数据的采集和分发，应包含：表输入、表输出、模型输入、模型输出、文件输入、文件输出、sql 输入等；转换组件用于完成数据的清洗转换，包含如下操作：聚合、集合、关联、路由、表达式、过滤、排序、清洗等；转换包含：一致性维度、行列转换等；脚本组件提供脚本灵活调用，支持：sql 脚本、shell 脚本、存储过程等方式。

5.3.8 清洗转换规则

产品支持清洗转换规则的封装，并提供扩展接口，用户只需要简单的配置即可完成数据的清洗转换。清洗转换规则应支持数据行级清洗和字段级清洗，行级清洗包含：行记录过滤、去重、抽样等，字段级清洗包含：字段的拆分、分割、合并、替换、格式转换等，清洗转换规则能根据配置规则项自动合并运算，批量处理大幅提升性能。

5.3.9 数据校验

提供对交换前后数据对比校验功能，交换过程边设计边对比，从数据层面来验证处理逻辑的准确性。

5.4 调度管理

5.4.1 交换任务流程设计器

交换任务流程设计器提供图形化的方式拖拽交换过程和流程组件，完成作业

的串行或并行设计。设计器支持放大、缩小、复制、语法检查、依赖配置、作业互斥配置等便捷设计功能。

5.4.2 交换任务组件支持

任务编辑器应除了基本的图形连线外，还提供丰富的组件来完成流程的控制，应包含：引用作业、分支、sql 脚本、shell 脚本、Email、文件监听等组件。

5.4.3 交换任务流程发布

为保证系统稳定运行的同时，开发任务也能执行，系统应从逻辑上划分为设计区和运行区，设计区设置的作业和交换任务流程能通过发布接口发布到运行区，完成开发的同时不干扰系统运行。

5.4.4 任务调度服务

具有灵活的调度服务功能，能基于时间窗口、事件、文件触发等多种方式来启动调度，调度的时间粒度支持分钟级别调用。

5.4.5 交换任务导入/导出

支持交换任务调度程序的导入/导出，能很方便的完成环境的移植和部署，导出的格式为标准的 xml 格式，导入时能根据名称和内容进行重复性判断，有选择的覆盖。

5.5 系统监控

5.5.1 任务运行监控

软件提供多种形式的运行状态的监控，支持图形监控和列表监控，图形监控提供任务运行的总览图，能直接的展现各交换任务的执行状态和执行时长信息，

列表监控提供明细的运行监控，粒度细化到作业内部的组件单元。

5.5.2 系统运行监控

产品提供了详细的系统运营情况监控报告，针对系统运行状态、操作耗时情况，资源占用情况、数据表存储情况、错误问题情况等多个维度全方面的监测系统的运营情况，提供的是实时统计数据，监控报告支持由汇总数据到明细记录的钻取等。

5.6 系统管理

5.6.1 备份与恢复

软件应提供数据备份的功能，防止系统因意外中断而发生数据丢失，数据备份内容应包括所有作业的配置、模型配置、交换任务配置等，支持完整或部分备份，同时对备份的文件进行恢复的时候可以进行选择性的恢复操作。

5.6.2 系统审计日志

软件支持日志审计功能，可涵盖系统所有的操作，如用户的登录注销、服务器的维护、用户管理和权限分配、模型变更、作业变更、作业发布、系统异常等。同时还应提供专门的日志查询和清理入口，方便的监控软件运行情况。

6 产品功能

6.1 参数配置

6.1.1 采集参数

针对全量采集和增量采集，可分别设置采集参数，针对读端、写端、运行三

个方面进行设置，可指定全字段获取或增量获取机制，批量加载机制、断点续传机制等模式。

6.1.2 分发参数

对数据传输时是否启用离线接收、传输超时天数，发送线程池的大小、线程空闲时保留时间，数据发送、接收备份目录、任务发送间隔、任务存储目录、任务备份目录，以及重发机制等设置，对分发任务各环节进行预置参数控制，并考虑离线和重发等情况。

6.1.3 监控参数

对传输交换管理平台的监控主要是对数据交换调度执行情况的监控，针对监控时点的当前任务数量、当前任务运行时间、今日完成任务数量、今日剩余任务数量、今日异常任务数。



图2 监控参数

6.1.4 全局参数

可设置传输交换管理平台全局参数，包括接收数据优先级、默认加密方式、

启动时是否进行接收数据检查、是否列出恢复的任务、AES 加密密钥、客户端认证信息、平台密级标识、推送周期、接收最大数量、完成任务保留天数、任务优先级等，用于总体控制平台执行和控制情况，保障平台稳定运行。

6.2 数据处理

6.2.1 作业设计器

传输交换管理平台提供了可视化的数据交换作业定义功能，支持作业的试运行和断点调试等，丰富的组件支持，能完成边调试作业边预览数据。

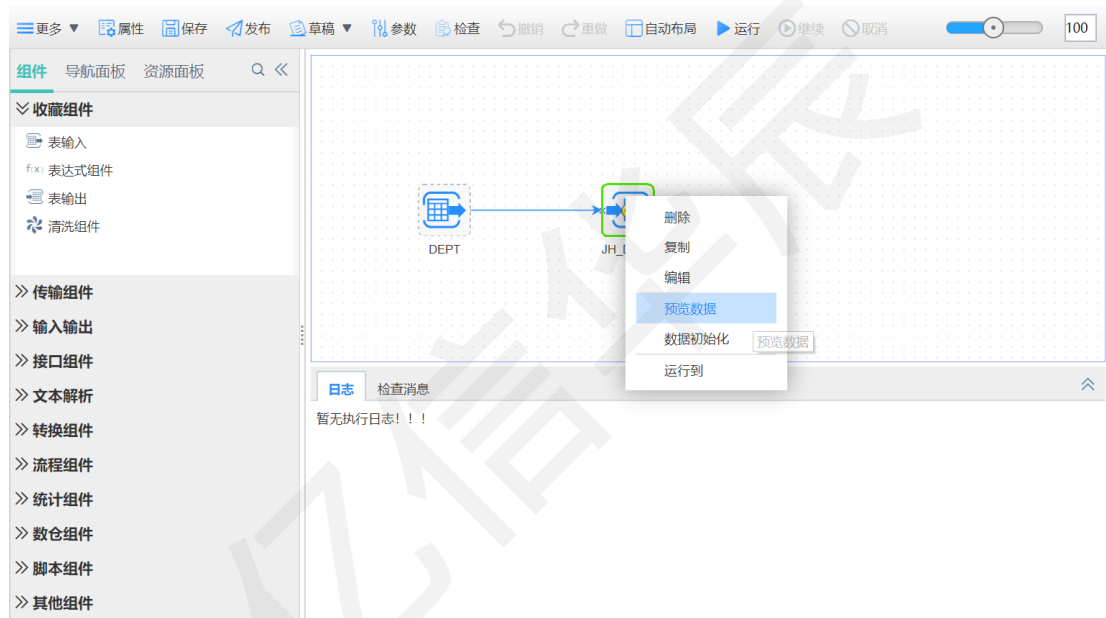


图3 作业设计器

6.2.2 输入输出组件

传输交换管理平台支持常用输入输出组件包括：



图4 输入输出组件

6.2.3 转换组件

另外，产品还内预置多种数据转换方式，可支持常用的聚合、集合、连接、过滤、排序、路由、表达式组件等，满足多种数据灵活处理场景需要。



图5 转换组件

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/045112221102011140>