****

**TongETL V24.2.6**

**产品功能清单**

**北京东方通科技股份有限公司**

**2024年**

**文档版本：V1.0**

**发布日期：2024年06月**

**版权声明**

**本产品功能清单的内容是东方通数据中台中TongETL产品的功能清单说明。本材料的相关权力归北京东方通科技股份有限公司所有。非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。**

**北京东方通科技股份有限公司**

**本资料将不定期更新，如欲获取最新相关信息，请访问东方通公司网站：[http：//www.tongtech.com/](http://www.tongtech.com/)**

**地址：北京市海淀区中关村南大街2号数码大厦A座22层**

**邮 编：100080**

**电 话：010-64135099**

目录

[1 ETL工具软件（TongETL） 1](#_Toc1007)

[1.1 架构设计 1](#_Toc30818)

[1.1.1 功能架构 1](#_Toc11446)

[1.1.2 技术架构 3](#_Toc14173)

[1.2 功能设计 3](#_Toc19059)

[1.2.1 功能清单 3](#_Toc14912)

[1.2.2 ETL服务器 5](#_Toc26908)

[1.2.3 管理中心 7](#_Toc13684)

[1.2.4 流程监控 13](#_Toc2709)

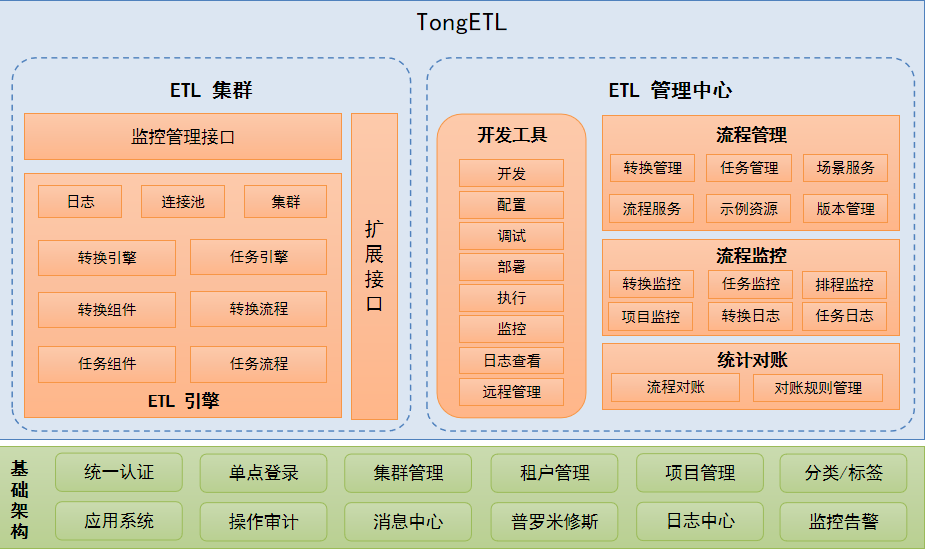
[1.2.5 日志审计 16](#_Toc19939)

[1.2.6 交换对账 17](#_Toc10328)

# **ETL工具软件（TongETL）**

## **架构设计**

### **功能架构**

­­­­­­

#### **服务器**

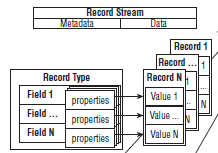
TongETL服务器是一个TongETL的逻辑节点，包含对数据集成流程的执行能力，并通过HTTP协议对外开放对服务器的管理接口。

* **ETL引擎**

TongETL服务器包含两个执行引擎：任务引擎和转换引擎，分别实现对任务调度管理的任务流程和完成对实际数据抽取、加工处理、加载的转换流程的执行。

在TongETL服务器启动时，可以检查部署在该节点上的数据集成流程，并执行配置为“自动重启动”的流程，其他非自动重启动的流程需要手工启动。

在TongETL数据流程中，数据是以行为单位进行处理，如下图所示，TongETL通过两个对象描述一个数据行：数据对象和元信息对象。数据对象实际的行数据，由一个或多个数据列（字段）组成；元信息对象用于描述数据对象，包括数据对象中每个列（字段）的类型、长度、格式等。



* **监控管理接口**

TongETL服务器还提供了对运行时的监控和管理功能，这些功能通过Java、HTTP、REST等多种封装形式对外开放。通过这些接口，第三方应用可以将监控管理功能集成到自己的管理控制台中进行统一监控管理，实现对TongETL的远程管理，如流程部署、流程控制、转换监控、日志查看等功能。

* **认证授权**

TongETL服务器提供了完善的认证授权机制，只有通过认证和授权的用户，才可以通过监控管理接口实现对服务器的管理和监控。

#### **管理中心**

TongETL管理中心集开发、调试、配置、部署、执行、监控、日志、管理等功能于一体的平台，是流程管理、监控和运维的一站式管理平台。

TongETL任务转换管理和监控能够全方位的管理所有流程，实现流程的全生命周期管理和自动化的版本管理，通过流程授权机制实现不同用户对流程的共享和权限控制，从而在权限控制之下实现流程的多人共享协作。

* **开发/调试**

管理中心提供大量的任务组件和转化组件，通过这些组件，以图形化的方式，实现数据集成流程的快速编排。

管理中心还提供了功能强大的调试预览功能，可以在开发过程中实现数据行级别的调试和预览，跟踪和观察每一行数据经过数据集成流程中每一个转化组件进行加工处理的结果。

* **远程管理**

通过管理中心和服务器的远程接口，实现对服务器的管理。包括数据集成流程的分布式部署、远程执行、对执行状态的实时监控、对执行日志进行查看和分析。

可以对运行中的流程执行暂停、开始、停止、解部署等控制，同时还可以对数据处理状态进行实时监控，包括每个组件处理的记录数、过滤的记录数，并且可以得到每个组件处理数据的性能指标和整个集成流程的性能指标。

支持全方位的监控能力，包括流程的实时运行情况、实时统计情况、流程各组件的耗时情况和异常情况、任务的计划排期和执行情况。

### **技术架构**

图示

描述已自动生成

## **功能设计**

### **功能清单**

|  |  |  |
| --- | --- | --- |
| 一级模块 | 二级功能 | 三级功能 |
| 系统概览 |  |  |
|  | 产品介绍 |  |
|  | 操作引导 |  |
|  | 业务统计摘要 |  |
| 分类管理 |  |  |
|  | 分类管理 |  |
|  |  | 分类增删查、重命名。 |
|  |  | 批量移动流程所属分类。 |
| 转换管理 |  |  |
|  | 流程列表管理 |  |
|  |  | 按分类查看 |
|  |  | 按数据源查看 |
|  |  | 流程查询、过滤、排序 |
|  |  | 分区管理 |
|  | 流程管理 |  |
|  |  | 创建流程 |
|  |  | 编辑流程 |
|  |  | 查看流程 |
|  |  | 删除流程 |
|  |  | 复制流程 |
|  |  | 流程检入检出 |
|  |  | 流程部署卸载 |
|  |  | 流程批量导入导出 |
|  |  | 流程配置导出 |
|  | 流程转换监控 |  |
|  |  | 转换启动停止 |
|  |  | 单次执行 |
|  |  | 补采 |
|  |  | 调度策略 |
|  |  | 查看转换详情 |
|  |  | 查看日志 |
|  |  | 单次执行监控 |
|  |  | 补采监控 |
| 任务管理 |  |  |
|  | 任务管理 | 同转换管理。 |
|  | 任务监控 | 同转换监控。 |
| 场景服务 |  |  |
|  | 场景类型 |  |
|  |  | 异构表全量同步 |
|  |  | 异构表触发器增量同步 |
|  |  | 异构表时间戳增量同步 |
|  |  | 同构表全量同步 |
|  |  | 同构表全库备份 |
|  |  | 同构表触发器增量同步 |
|  |  | 同构表时间戳增量同步 |
|  | 场景服务管理 |  |
|  |  | 创建场景服务 |
|  |  | 查看场景服务 |
|  |  | 编辑场景服务 |
|  |  | 删除场景服务 |
| 流程服务 |  |  |
|  | 流程服务实例管理 |  |
|  |  | 创建实例 |
|  |  | 查看实例 |
|  |  | 实例监控 |
|  | 系统流程服务管理 |  |
|  |  | 创建服务 |
|  |  | 查看服务 |
|  |  | 服务监控 |
| 示例资源管理 |  |  |
|  | 示例资源分类 |  |
|  |  | 增量同步 |
|  |  | 全量同步 |
|  |  | 数据处理组件 |
|  |  | 文件组件 |
|  |  | 工具组件 |
|  |  | 流程控制组件 |
| 流程监控 |  |  |
|  | 排程监控 | 查看部署流程的排期监控，能够直观显示未来一段时间内的任务排期情况，可细化到具体每天某个时段的任务排期计划。 |
|  | 转换监控 | 支持查看和管理用户有权限查看的转换流程。 |
|  | 任务监控 | 支持查看和管理用户有权限查看的任务流程。 |
|  | 项目监控 |  |
|  | 转换统计监控 | 查看项目下转换流程的执行状况统计。 |
|  | 任务统计监控 | 查看项目下任务流程的执行状况统计。 |
|  | 连接池监控 | 查看运行态集群使用的数据源连接池状况。 |
| 日志审计 |  |  |
|  | 转换执行日志 | 查看并搜索转换流程的执行日志。 |
|  | 任务执行日志 | 查看并搜索任务流程的执行日志。 |
|  | 流程组件日志 | 查看转换流程的各个组件的详细执行日志和统计结果。 |
|  | 流程调度日志 | 查看流程的调度批次日志。 |
|  | 结果文件日志 | 查看转换流程的处理过的文件的日志详情。 |
| 交换对账 |  |  |
|  | 流程对账 | 支持查看流程对账结果。 |
|  | 对账规则 | 支持基于自定义对账规则的转换执行结果的对账。 |

### **ETL服务器**

* **ETL引擎**

TongETL服务器包含两个执行引擎：任务引擎和转换引擎，分别实现对任务调度管理的任务流程和完成对实际数据抽取、加工处理、加载的转换流程的执行。

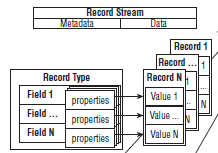
* **数据处理**

TongETL采用异步并行处理的方式实现数据的高效处理，数据在TongETL中类似于流水线上的产品，逐行流经流程中的每个组件，每经过一个组件就被加工成一个既定格式的中间状态。数据经过一个组件的处理后被迅速交给下一个组件处理，同时当前的组件已经开始处理新的数据。

TongETL提供集群部署方式，允许将转换或转换中的比较耗时的数据处理组件部署在多台服务器上并发执行，从而将转换的工作分摊到多台服务器上，提高TongETL的数据处理效率。

在TongETL服务器启动时，可以检查部署在该节点上的数据集成流程，并执行配置为“自动重启动”的流程，其他非自动重启动的流程需要手工启动。

在TongETL数据流程中，数据是以行为单位进行处理，如下图所示，TongETL通过两个对象描述一个数据行：数据对象和元信息对象。数据对象实际的行数据，由一个或多个数据列（字段）组成；元信息对象用于描述数据对象，包括数据对象中每个列（字段）的类型、长度、格式等。



* **监控管理接口**

TongETL服务器还提供了对运行时的监控和管理功能，这些功能通过Java、HTTP、REST等多种封装形式对外开放。通过这些接口，第三方应用可以将监控管理功能集成到自己的管理控制台中进行统一监控管理，实现对TongETL的远程管理，如流程部署、流程控制、转换监控、日志查看等功能。

* **适配器**

基于标准的JDBC、ODBC接口，实现对各种主流数据库系统的支持。如Oracle、DB2、SQL Server、Sybase、InfoMix、vertica、doris等主流数据库，MySQL、PostgreSQL等开源数据库，达梦、瀚高、神通、GBase8t、KingBase、高斯、昆仑等国产数据库。

提供丰富的数据文件抽取和加载组件，支持包括普通文本、CSV、XML、Excel等多种格式的文件。

支持HTTP、JMS、FTP、Web Services等协议和其他应用系统进行交互。

### **管理中心**

#### **开发工具**

通过TongETL 内置的大量组件和B/S的可视化图形界面，使用人员无需编码，既可以灵活、方便地定制出各种数据集成流程，又能够减少手工代码的错误，还可以利用调试预览及时发现潜在错误，提升集成质量，并能对ETL服务器和资源库进行管理，全面提高集成速度。

#### **开发/调试**

TongETL提供大量的任务组件和转化组件，通过这些组件，以图形化的方式，实现数据集成流程的快速编排。

TongETL还提供了功能强大的调试预览功能，可以在开发过程中实现数据行级别的调试和预览，跟踪和观察每一行数据经过数据集成流程中每一个转化组件进行加工处理的结果。

通过集成开发工具开发调试完成的数据集成流程统一存储管理。

#### **远程管理**

TongETL实现对部署流程的远程管理。包括数据集成流程的分布式部署、远程执行、对执行状态的实时监控、对执行日志进行查看和分析。

TongETL的监控管理功能可以对运行中的流程执行暂停、开始、停止、解部署等控制，同时还可以对数据处理状态进行实时监控，包括每个组件处理的记录数、过滤的记录数，并且可以得到每个组件处理数据的性能指标和整个集成流程的性能指标。

#### **处理组件**

TongETL内置大量的任务组件和转换组件，用户可以通过拖拽的方式快速完成各种复杂数据集成需求和集成的调度控制。提供的转换组件覆盖数据映射、数据丰富、数据计算、数据验证、数据排序、数据合并、数据拆分、数据生成、数据去重、数据分组、行列转换等复杂处理，提供的任务组件涵盖定时调度、周期循环调度等调度模式组件、以及数据处理的一些前置、后置检查操作等。

#### **概览**

概览页面展示产品介绍内容、产品操作引导和业务统计摘要内容。

#### **转换流程**

转换流程负责完成数据集成过程中的数据抽取、转换、加载工作，转换流程由多个转换组件编排而成。转换组件是转换流程的最小处理单元，每个转换组件完成一个特殊的数据处理任务，多个转换组件组成一个转换流程。正是因为TongETL有丰富的转换组件，使得TongETL具备高水准的数据加工处理能力。

转换组件整合了DataX，实现了包括 MySQL、Oracle、OceanBase、SqlServer、Postgre、HDFS、Hive、ADS、HBase、TableStore(OTS)、MaxCompute(ODPS)、Hologres、DRDS 等各种异构数据源之间高效的数据同步功能。

TongETL提供的转换组件如下图所示：



#### **任务流程**

任务流程负责整个数据集成过程的调度和管理，如定时执行、周期执行、告警等。任务流程一般由转换流程组成，可以对多个转换流程和其他任务流程（子流程）进行流程编排。任务组件就是任务流程的最小处理单元，如开始组件负责启动任务流程，可以设定任务流程的执行方式，如一次性任务，定时启动的任务，循环性的任务等。

TongETL提供的任务组件如下图所示：



#### **任务/转换管理**

1、 新增任务

任务新增时，增加任务创建者，所有任务全部为本人私有。

2、 修改任务

对于自己的任务，在任务没有部署前的修改工作

4、 任务检出

当多人共享的服务，在进行修改前，需要将任务捡出，此时任务被锁定，在此任务没有被检入之前，其他用户只能查看，不能修改

开发工具中打开任务，默认为检出，如果这个任务已经被检出，则提示只能查看，不能修改

5、 任务检入

将捡出的任务信息提交到资源库中，释放流程锁，允许其他用户继续检出编辑，检出用户关闭流程时，提示进行检入操作。若用户退出但流程仍未检入，也会继续保持检出状态，直至检出流程的用户，执行检入操作。

7、任务部署

将任务部署到运行节点上，所有的部署操作全部为本地部署或ETCD部署。

多人共享流程的时候，处于检出状态的流程不能部署。只能部署草案版本。无多人共享的流程，可以随时部署，不受限制。每次部署后，新增一个产品版本。

8、任务删除

只有owner才能删除，删除权限不能分享。无论该流程是否分享，也无论该流程是否被捡出，均可删除

9、任务复制

可以查看别人的任务，然后复制整个任务的配置信息，变成自己的私有任务。

将指定的流程复制一份，另存为一个新的流程，能查看就能复制。

10、任务版本

【草案版本】：流程最新的可编辑的版本是草案版本，第一次创建时以流程名称为key。草案版本可以编辑（需要有权限和先检出成功）、部署。

【产品版本】：流程部署后的流程快照，产品版本不可被修改，不可以被再次部署。可以覆盖此流程的草案版本， 或另存为一个新流程的草案版本。

11、任务版本回滚

将任务回滚到指定的版本，相当于更新草案版本

#### **场景服务**

场景服务将TongETL的部分常用能力抽象为一个个典型场景，用户无需关心流程各组件的具体配置和逻辑，仅需要从业务视角定义事件的相关参数既能够实现用户的业务场景。

场景服务目前内置了数据交换和接口数据采集类场景，如典型的同构库数据备份、异构库同构表的数据全量和增量同步、异构系统文件传输等等。

典型的场景服务有：

关系数据库交换场景

异构库同构表—时间戳模式场景服务

异构库同构表—触发器模式场景服务

异构库全库数据备份模式场景服务

异构库异构表—时间戳模式场景服务

异构库异构表—触发器模式场景服务

异构库异构表全量数据同步

#### **示例资源**

能够展示系统内置的任务demo和模板的示例文件，以便用户能够快速学习ETL的典型应用场景，支持用户对示例的查看和复制功能：

查看：用户点击示例名称后进入流程查看界面，用户可查看流程中各组件的配置信息，可对其修改但并不保存用户的修改。

复制：当用户通过查看功能发现此示例和自己所需要的流程相似，希望能够通过示例流程快速完成自己流程开发时，可复制当前示例然后进行后续的编辑、测试等操作。

#### 流程服务

流程服务将TongETL转换流程能力封装为服务的形式供第三方系统调用，提供了一种ETL核心能力对外开放的技术方式。用户可创建系统流程服务，形成一个流程服务的“模版”，后续通过调用该服务，传入具体的流程配置参数，生成流程服务的实例，以高效实现外部系统调用TongETL进行同一类数据转换的需求。

TongETL提供流程服务能力，用户可将转换流程部署为流程服务，通过外部API调用实现任务调度，系统自动为每一个服务调用生成一个服务实例，从而解决相同类型的转换仅需要通过变量即可实现业务逻辑，而无需重复定义多个转换流程。

流程服务通过将流程中可变环节进行参数化，实现用一个流程满足多个同类业务需求的场景，流程服务的典型场景为多表数据同步、文件同步、文件自动生成等，以将库表增量数据导出为文件的流程服务为例，其使用方式如下图所示：



用户仅需要开发并部署一个流程服务，每个业务客户端在请求流程服务时传入自己的业务参数，如图中所示的具体表名，流程服务在收到客户端请求后会自动创建一个示例并完成流程中规定的业务，流程运行结束后回收相应资源并结束运行实例。

### **流程监控**

#### **排程监控**

展示已部署流程的计划执行时间的统计，以便发现规律和查找问题，优化流程的部署。

支持根据时间区间、项目列表，筛选项目下已部署流程的计划执行时间。

支持按照日期展示已部署流程的计划执行时间统计。

支持按照时段展示已部署流程的计划执行时间统计。

支持展示已部署流程的计划执行时间明细数据。

#### **转换监控**

获取转换监控实时信息，控制转换流程的运行状况。

支持展示用户有权限查看的已部署的转换流程列表。

支持展示用户有权限查看的已部署流程的项目名称、流程名称、版本、部署的运行周期、部署的集群名称、当前启停状态、创建人、部署时间等详细信息。

支持展示流程的部署详情。

支持查看流程的运行日志。

支持已部署流程的单次执行。

支持启动、停止、卸载、删除已部署的流程。

支持对已部署的流程进行补采操作。

支持修改已部署流程的调度周期。

支持查看已部署流程的单次执行状况监控、补采批次运行状况监控。

#### **任务监控**

获取任务监控实时信息，控制任务流程的运行状况。

支持展示用户有权限查看的已部署的任务流程列表。

支持展示用户有权限查看的已部署流程的项目名称、流程名称、版本、部署的运行周期、部署的集群名称、当前启停状态、创建人、部署时间等详细信息。

支持展示流程的部署详情。

支持查看流程的运行日志。

支持已部署流程的单次执行。

支持启动、停止、卸载、删除已部署的流程。

支持对已部署的流程进行补采操作。

支持修改已部署流程的调度周期。

支持查看已部署流程的单次执行状况监控、补采批次运行状况监控。

#### **项目监控**

支持展示指定时间区间内的各个项目下流程的批次运行状态汇总信息。

支持展示指定时间区间内的运行的流程总数。

支持展示指定时间区间内的转换流程运行总批次数、成功/失败/停止/拒绝的批次数目。

支持展示指定时间区间内的任务流程运行总批次数、成功/失败/停止/拒绝的批次数目。

#### **转换统计监控**

支持展示指定时间区间、指定项目下的转换流程批次运行统计信息。

支持展示转换流程的运行集群、调度周期、成功/失败/停止/拒绝的批次数目、当前运行状态、创建人、创建时间等信息。

支持查看转换流程的部署详情。

支持查看转换流程的运行日志。

#### **任务统计监控**

支持展示指定时间区间、指定项目下的任务流程批次运行统计信息。

支持展示任务流程的运行集群、调度周期、成功/失败/停止/拒绝的批次数目、当前运行状态、创建人、创建时间等信息。

支持查看任务流程的部署详情。

支持查看任务流程的运行日志。

#### **连接池监控**

支持展示ETL集群各个运行态中关系型数据源连接池监控的实时信息。

支持展示数据源的名称、类型、初始连接数、IP、端口、最大连接数、当前活动、空闲连接数，最大空闲和最小空闲连接数信息。

支持对比运行态使用的数据源信息和元数据中的配置是否一致。

支持展示运行态使用的数据源信息和元数据中的配置不一样的配置项。

### **日志审计**

#### **转换执行日志**

支持展示转换流程的运行日志明细信息。

支持展示转换流程的各个运行批次的项目名称、流程id、批次id、转换名称、接收数据量、处理数据量、差异数据量、执行结果状态、运行类型、运行时长、开始时间、结束时间、日志时间、批次运行的所属服务器地址等信息。

支持查看批次运行的运行日志。

支持将批次立即重跑一次。

支持忽略此批次日志，忽略后将不展示、不统计此批次。

支持展示批次运行的参数列表。

#### **任务执行日志**

支持展示任务流程的运行日志明细信息。

支持展示任务流程的各个运行批次的项目名称、流程id、批次id、任务名称、接收数据量、处理数据量、差异数据量、执行结果状态、运行类型、运行时长、开始时间、结束时间、日志时间、批次运行的所属服务器地址等信息。

支持查看批次运行的运行日志。

支持将批次立即重跑一次。

支持忽略此批次日志，忽略后将不展示、不统计此批次。

支持展示批次运行的参数列表。

#### **流程组件日志**

支持展示转换流程各个组件运行的详细日志信息。

支持展示各个组件运行的批次id、流程名称、组件名称和业务统计数量、元数据依赖信息、运行状态、服务器所在ip地址、运行时长、开始时间、结束时间等信息。

#### **结果文件日志**

支持展示转换流程的文件类组件处理过的文件的文件名称、文件行数、文件大小等详细信息。

#### **流程调度日志**

支持展示ETL流程的调度日志。

支持展示流程调度的调度时间、运行状态、开始时间、结束时间、耗时、重试次数等信息。

支持停止正在运行的调度批次。

支持批量重跑流程运行批次。

### **交换对账**

#### 交换对账配置

针对每个流程，配置每个流程的对账规则

获取到要配置流程的组件信息，包括副本信息

配置流程对账规则，支持+- sum （）

如组件1.输入+组件2.读取=组件3.写入+sum（组件4副本.输出）+组件5.异常

#### 对账结果展示

按对账规则所列的左右两侧作为对账数据，按照用户查询的时间范围进行展示，所有时间范围只控制任务开始时间，不控制结束时间。

按任务执行批次继续显示对账结果，找到不平的数据后继续追踪，显示原始统计信息。