**数据清洗服务**

**一、数据清洗服务**

1.1 低耦合

数据清洗服务是一款独立部署的产品，通过RESTful接口实现大数据平台与服务之间的交互。

1.2 高效率

数据清洗服务支持配置作业，用于定时执行任务。也可以在EDI中通过相应配置，在相关抽取结束后，自动开始执行清洗。

清洗过程高度自动化，无需人员干预，自动将问题数据提炼出来。

**二、清洗规则配置**

包括：

1.1 空值检查

空值检查用于检查关键字段非空。

1.2 值域检查

值域检查用于检查关键字段的取值范围，支持数值型、日期型字段检查。

1.3 规范检查

规范检查用于检查用于检查指标值的格式是否规范，支持身份证、统一社会信用代码等多种数据类型的检测，支持自定义正则表达式。

1.4 代码项检查

代码项检查用于检查性别、交易方式、招标方式等代码值或者文本内容是否规范。

1.5 高度个性化检查

服务提供相应接口，可以用于更加复杂的个性化配置。

**三、数据治理功能**

清洗服务提供了可视化的数据查看页面。

支持随时查看清洗结果。

支持导出异常数据报告。

支持在线修改问题数据，并及时同步大数据平台。

1. **异常数据报告**

支持按各地区分类。不同地区的异常数据归到不同的异常报告中。

支持交易平台链接。报告中每条数据都可以提供相关超链接，点击后直接查看交易平台中的相关数据。

1. **流程设计**

大数据平台

数据清洗服务

公共资源交易平台

公共服务平台

正常数据

临时表

统计表

基础表

异常数据报告反馈