

BEYOND
TECHNOLOGY

SaCa[®] DataTransform

数据转换与清洗平台软件

产品白皮书

目录

一、前言	3
二、产品简介	4
三、系统架构	5
四、产品功能	6
4.1 异构数据源整合	6
4.2 图形化映射工具	6
4.3 多样化数据转换	6
4.4 分布式并行运行	6
4.5 异常自动恢复	6
4.6 Web 方式管理监控	6
五、产品优势	7
5.1 卓越的性能优势	7
5.2 敏捷的配置工具	7
5.3 开箱即用的模板	7
5.4 自动化异常修复	7
六、总结	8
七、典型应用项目	9
八、关于东软	10

一、前言

现在，企业所烦恼的不再是数据匮乏，而是无法在正确的时间获得正确的数据。政府机构面临的挑战不再是信息的孤岛，而是对各类数据进行整合。

由于组织结构和管理模式等方面的原因，不同职能部门间在过去都是分别独立建立系统，缺乏统一的协调和规划，往往只从具体业务出发，实现手工作业到电子化的转变，很少考虑系统间的信息的整合与共享，导致系统数量多，异构化，数据量大、复杂性大，数据的分散、冗余、不完整。

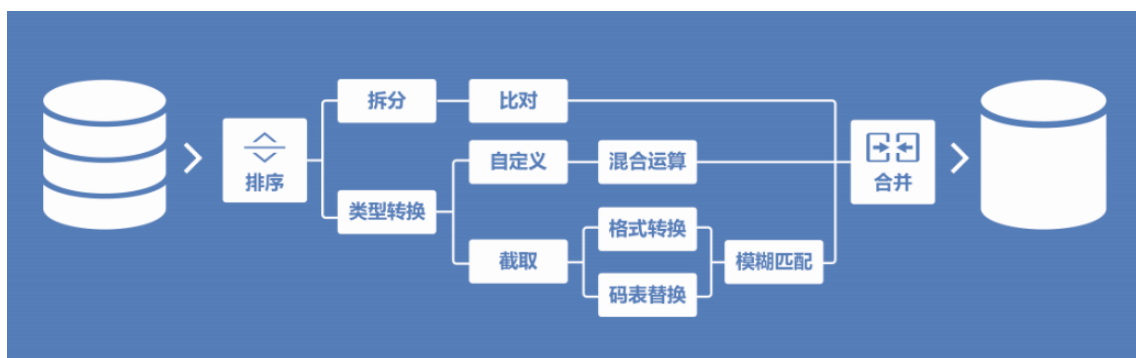
数据整合的核心内容是数据的标准化处理，也称为数据的转换与清洗，从数据源中抽取数据，然后对这些数据进行转化，最终加载的目标数据库或者数据仓库中去，这也就是我们通常所说的 ETL 过程。在 SaCa Data Integration 数据集成工具套件中，SaCa DataTransform 为整个 ETL 过程提供了一个完整的设计与运行环境。



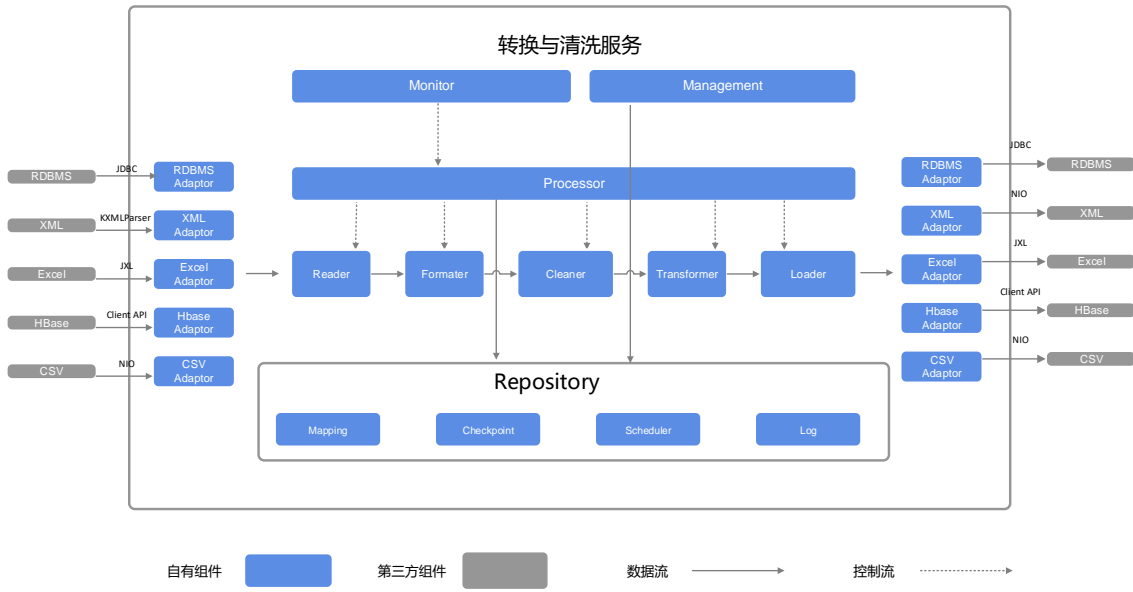
二、 产品简介

SaCa DataTransform 为具备高可扩展性与高性能的企业数据集成软件确立了标准。通过单一方式访问、转换、清洗和加载数据，适用于数据仓库、数据迁移等应用场景。该软件的高可扩展性可以支持海量数据并满足企业对安全和性能的需求。

内置大量数据转换模型，遵从国际化标准以及国家众多行业标准，几十种通用的转换规则如类型转换、字段拆分、字段合并、字符串处理、日期转换、算术运算、码表转换等，同时提供自定义转换接口实现特殊的数据转换处理。



三、系统架构



四、 产品功能

4.1 异构数据源整合

支持主流的关系型数据库 (Oracle/DB2/SQLServer/MySQL 等)、XML 文档、文本数据、Excel、Hbase、hive、impala 以及 Webservice 等接口, 支持灵活的扩展新的数据源

4.2 图形化映射工具

提供创建数据转换流程的图形化配置工具, 使数据集成配置人员能够快捷地定义数据转换流程中数据抽取 (Extract)、转换 (Transform)、加载 (Load) 过程。

4.3 多样化数据转换

如类型转换、字段拆分、字段合并、字符串处理、日期转换、算术运算、码表转换等, 同时提供自定义转换接口实现特殊的数据转换处理。

4.4 分布式并行运行

支持多个无关联任务并行执行, 并且可以在监控页面统一进行任务运行状态的查看、管理

4.5 异常自动恢复

通过日志、事务、故障处理等手段, 保证系统运行的可靠性, 以及业务过程的安全性, 出现错误时, 可以恢复运行, 支持断点续传

4.6 Web 方式管理监控

提供完全基于浏览器方式的图形化任务流程配置和监控, 更直观的展现数据流转情况, 无需安装客户端, 可以在任意可连同服务器的网络环境下配置和使用。

五、 产品优势

5.1 卓越的性能优势

利用分布式多机并行处理技术，将单个数据转换清洗任务并行化处理，提升整体数据处理效能，每秒可处理单表数据量>10 万条

5.2 敏捷的配置工具

图形化拖拽方式配置数据清洗过程，无论是单个业务还是大批量业务都可以快速完成，提升配置与管理人员的生产能力

5.3 开箱即用的模板

内置大量数据转换模型，遵从国际化标准以及国家众多行业标准，几十种通用的转换规则，降低二次开发成本。

5.4 自动化异常修复

提供网络中断、意外停机的错误处理能力，内置错误处理策略，配合断点续传、断点续抽等功能可按照预定规则进行自动化处理。

六、 总结

在中国企业信息化进程不断深入和数据市场快速发展的进程中，SaCa DataTransform 经过数年多个版本的研发和磨练已经形成了一套系统、成熟的产品体系结构。通过对中国特有的数据环境复杂性进行深入研究的基础上提出了适合中国市场的解决方案，成功在政府部门、企事业单位等不同行业不同应用的数据项目构建中应用。

七、 典型应用项目

行业	典型案例
政府	青岛市财源信息管理系统开发
	西安地铁线网（应急）指挥中心
	国家工商行政管理总局数据中心
	潍坊滨海经济技术开发区综治信息系统
	天山云容灾系统建设平台产品采购
	天水市公安局智慧公安信息化建设项目
	陕西省住房保障项目
其他	双胞胎（集团）股份有限公司
	中山大学附属第三医院
	重庆麦森云科技有限公司
	食品溯源公共平台
	四川应急指挥系统
	江苏省检察院案件管理综合查询系统

八、 关于东软

东软创立于 1991 年，是中国领先的 IT 解决方案和服务提供商。公司主营业务覆盖软件产品与平台、行业解决方案、产品工程解决方案和服务四个领域。目前，东软拥有员工 13000 余名，在中国建立了 8 个区域总部、16 个软件开发与技术支持中心，5 个软件研发基地，在 40 多个城市建立营销与服务网络，在大连、南海、成都和沈阳建立了 3 所东软信息学院和 1 所生物医学与信息工程学院；在美国、日本、香港、阿联酋、匈牙利和印度设有子公司。

东软是中国第一家上市的软件企业，是第一家通过 CMM5 和 CMMI (V1.2) 5 级认证的软件企业，是中国最大的离岸软件外包提供商。2007 年，公司主营业务收入为 33.5 亿元人民币。

东软的 IT 解决方案广泛地应用于电信、电力、社保、金融、税务、交通、教育、医疗、制造业以及电子政务等几十个重点行业和领域，在中国市场，拥有客户达 15000 家，其中在社保行业占有 50% 以上的市场份额，在电信行业占有 30% 的市场份额，在电力行业占有 10% 的市场份额，在网络安全领域拥有 15% 以上的市场份额。在离岸软件外包方面，东软已经与日本、美国、芬兰、荷兰、德国等国家的跨国企业建立战略合作伙伴关系，拥有 50 多家国际软件外包客户。2007 年，东软被美国国际外包专业委员会 (IAOP) 评为全球 25 家最优秀的外包提供商之一。

东软致力于成为最受社会、客户、投资者和员工尊敬的公司，并通过过程与方法的不断改进，领导力与员工竞争力的发展，持续和开放的创新，使公司成为全球优秀的 IT 解决方案和服务提供者。

产品网站: <http://platform.neusoft.com>

技术社区: <http://plus.neusoft.com>

电话: 400 655 6789

邮箱: platform@neusoft.com

微信: 东软平台产品

