

BEYOND
TECHNOLOGY

SaCa[®] Data Integration

数据集成套件

产品白皮书

目录

一、	前言	3
二、	关于 SaCa Data Integration	4
三、	一体化方案.....	5
1.	数据采集	6
2.	页面采集	7
3.	数据交换	8
4.	转换清洗	9
5.	数据比对	10
6.	数据质量	11
7.	数据服务	13
8.	数据目录	14
9.	元数据管理.....	16
四、	总结	18
五、	东软集团简介	18

一、前言

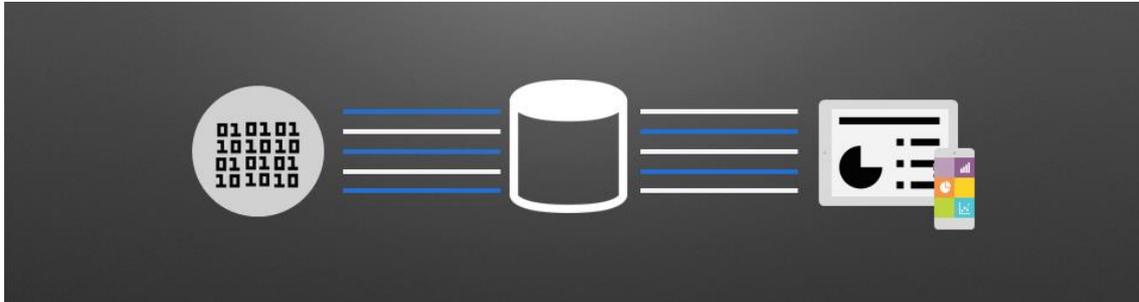
随着企业对业务过程处理的卓越追求和对市场需要的及时响应，出现了一系列的以“快速满足需求”为中心的业务系统。这些业务系统在一定程度上满足了当时企业发展的需要，但同时也对企业未来的发展埋下了一定的隐患。这些以“快速满足需求”为中心的业务系统，由于其业务特点的各异性，往往采用了不同的技术实现、系统和数据规范，这不仅增加了企业信息化环境的复杂性、业务信息的不对称和不一致性、业务信息的分散性和无全局性，还增加了运行和维护部门的负担，最终还分散了企业关注支持新业务需求的精力。

无论是政府还是各个行业的企业，都已经深深的意识到了这些隐患，从全局规划角度出发各种类型的标准化和流程化工作应运而生，并且被提到了企业发展的战略高度。政府部门正不断实施各种信息采集、整理，建立公共数据开放机制，促进政府部门间的数据交换，同时向社会开放数据，推动数据的开放和共享，而数据集成作为这些工作的基础设施，将起到支撑的作用。

二、 关于 SaCa Data Integration

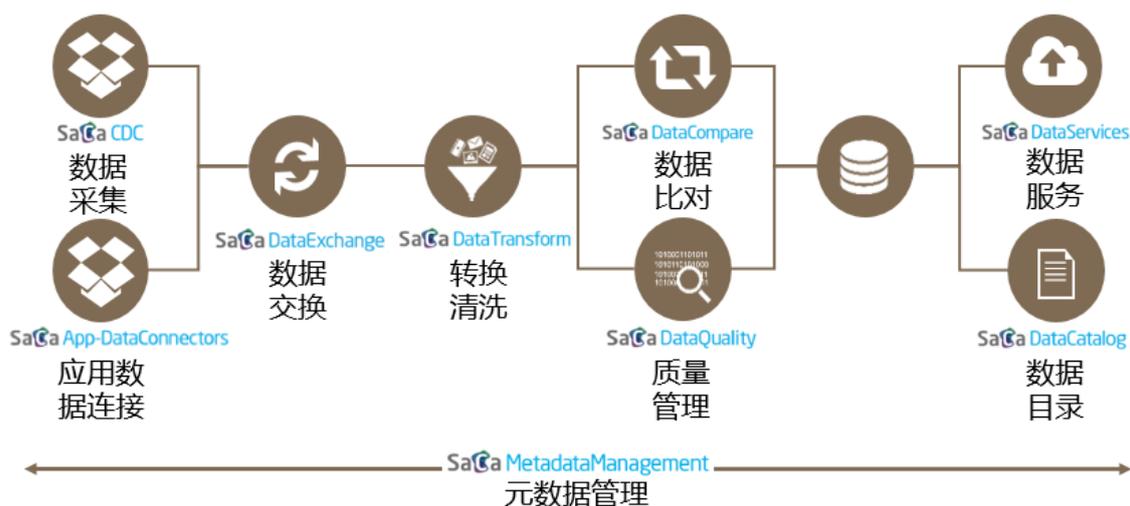
SaCa Data Integration 数据集成套件(简称 SaCa DI 套件), 是一系列数据整合产品集合, 涵盖了元数据管理、数据标准管理、数据实时采集、数据交换传输、数据质量管理、主数据管理以及数据开放与共享管理等多个方面。

这些产品支持使用系统的方法来解决政府部门及企业如今所面临的信息整合难题, 提供一体化解决方案, 用来透明地管理当今企业各类系统中庞杂的数据, 支持各种类型的业务系统数据的整合、交换和共享, 从而形成一套准确、干净、完整的数据集合。



三、 一体化方案

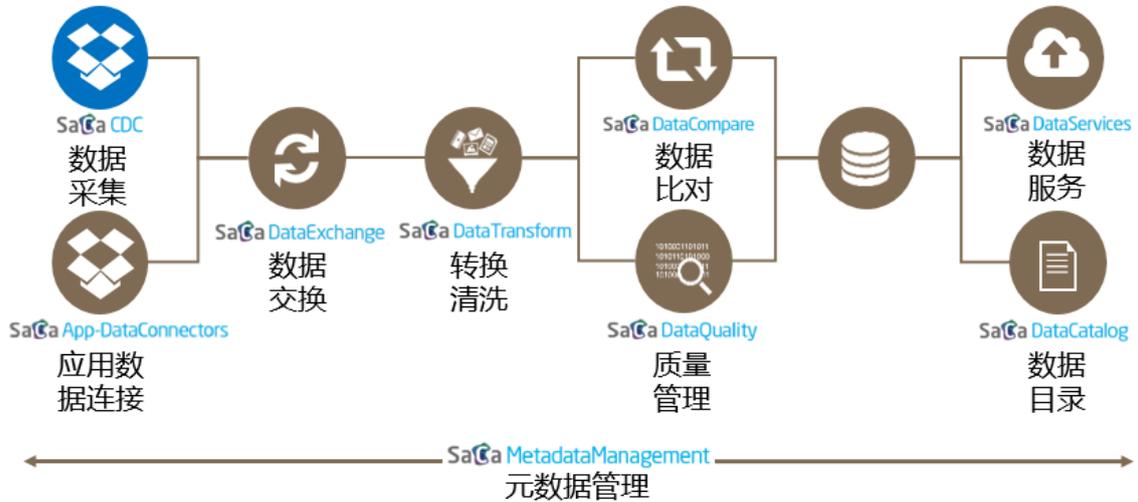
SaCa DI 套件一直致力于为客户提供具有强大的实时数据采集、数据交换，数据转换清洗、数据质量管理、数据比对和统一数据服务平台。



拥有遍布全球的客户，客户范围涵盖金融、电信、制造、政府、能源、保险、公众服务等多种行业。

借助 SaCa DI 套件，可以透明地处理当今企业各类系统中庞杂的数据，建立统一数据交换和共享平台；助力企业数据标准化和企业统一客户视图的建设工作，为建立以客户为中心，以服务为导向的新型服务企业提供底层信息流动支撑；提供对历史系统数据、私有系统数据的访问，充分利用企业信息化进程积累的历史数据；降低数据冗余度、降低数据不一致性和提高数据准确性，提高作为企业战略资源的各种信息和历史数据的使用率和使用准确率；建立统一数据服务平台，支持现有和未来业务系统的数据服务；助力企业实现跨地域的、跨业务系统的客户一致化服务水平，提高企业客户满意度和企业竞争能力。

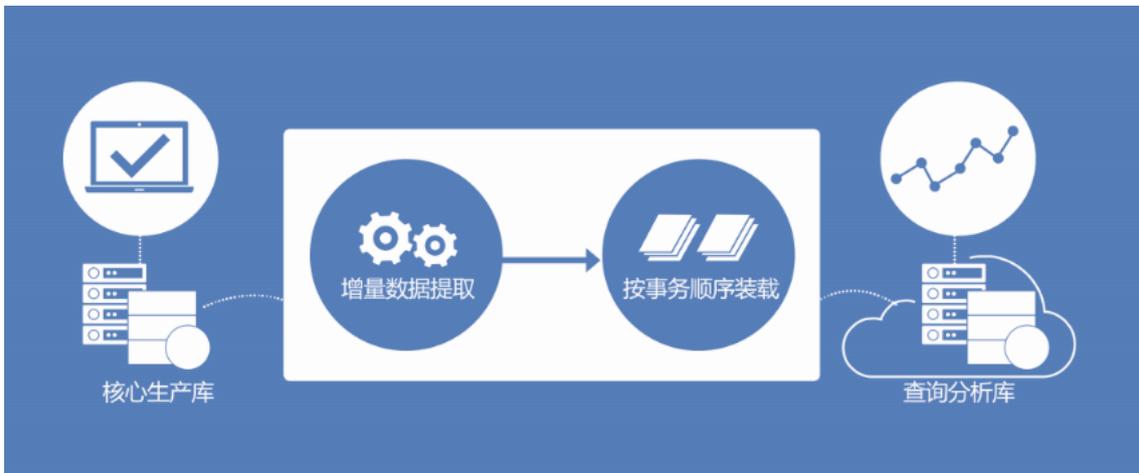
1. 数据采集



数据采集是所有数据系统必不可少的，随着大数据越来越被重视，数据采集的挑战也变的尤为突出。这其中包括：

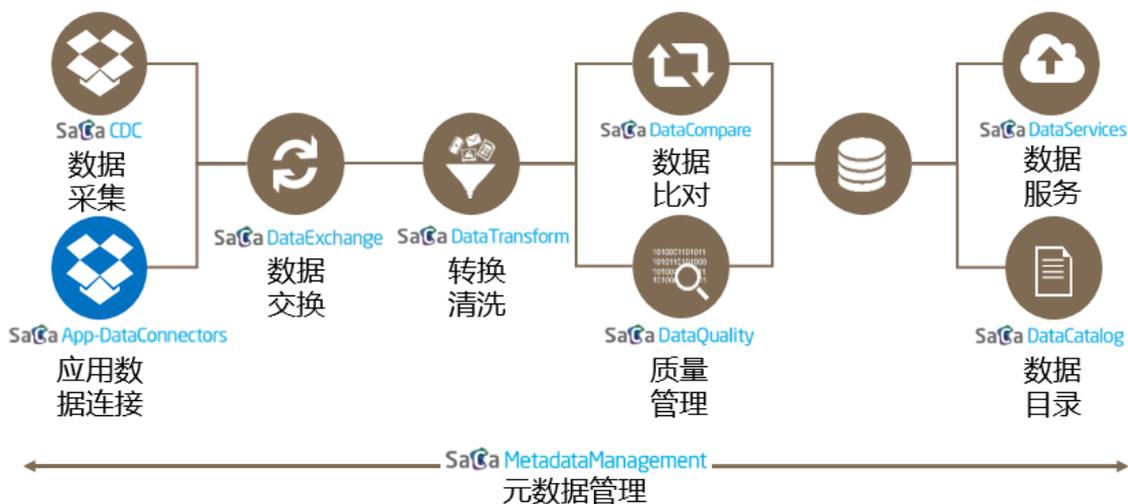
数据源多种多样、数据量大、变化快。如何保证数据采集的可靠性的性能，数据采集的数据源连接能力是非常重要的，这将直接决定它能够应用的范围。

东软 SaCa CDC 变更数据捕获平台，在 SaCa DI 套件中提供数据采集的能力。



SaCa CDC 基于日志采集与分析技术，近乎实时的捕获和保存变更数据，以独立于创建信息应用的方式，最大程度地减少对原系统的影响，广泛应用于数据仓库实时数据供给、实时商务智能、数据备份与容灾、数据库升级和移植、双业务中心、跨异构数据源和操作系统实时数据交付等场景。

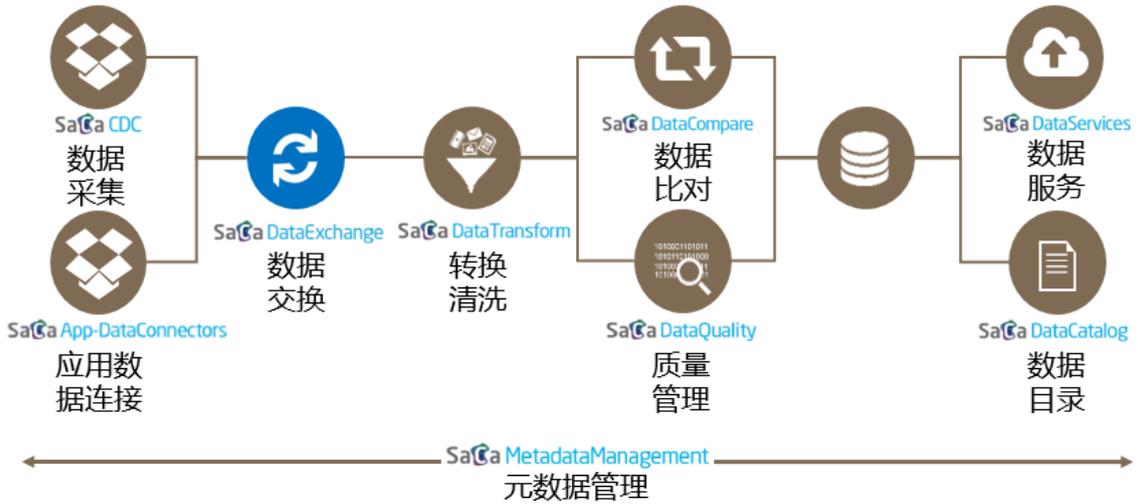
2. 页面采集



东软 SaCa App-DataConnectors 应用数据连接平台，在 SaCa DI 套件中提供页面数据的采集、分发能力。可以快速帮助用户实现信息一次录入，多系统同时保存的诉求，同时可以在不访问数据库和源代码的情况下，轻松获取僵尸系统的历史数据，以及实现老旧系统的移动端横向扩展开发。

SaCa App-DataConnectors 应用数据连接平台，是一款基于系统展现层与应用层的数据连接、开采与融合产品，能够智能地识别与连接不同业务系统之间的应用逻辑，通过重构已有系统的数据读写接口，实现异构系统之间的数据实时获取、业务融合、交互共享和无缝集成，从而实现应用层的系统开放，进而达到对已有应用系统的开放、对接、交互、重构等建设目标，形成跨平台、跨领域的采集、开放、共享、融合的大数据支撑环境。

3. 数据交换



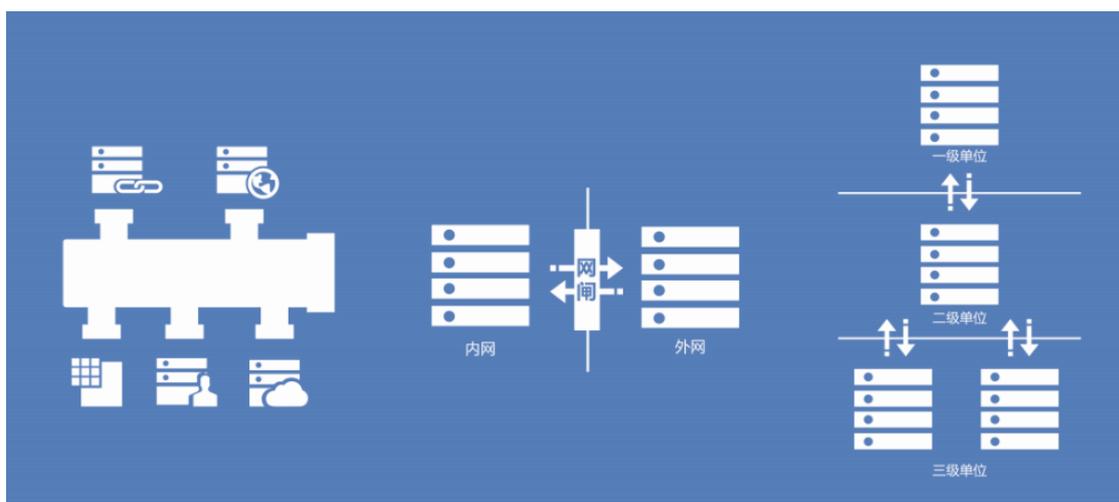
政府各部门电子政务平台的信息格式和交换方式的不一致，造成了政府各部门间难以形成有效的信息的交换和共享，政府各部门间电子政务平台成为了信息孤岛，影响了政府各部门间工作协同的实现，更达不到政务信息和政务服务的公开的目的。例如，在工程建设领域，由于各个地方之间的项目信息和信用信息的不共享，造成无法对工程建设领域的违规责任者进行全国联动监管，在工程项目审批和招投标的时候无法对此违规企业进行市场准入的限制，该企业仍可在其他地方继续进行投标甚至中标，并继续建设施工质量问题的工程，最终造成多起安全事故。

可见，信息的交换和共享对于电子政务平台整体的应用效果起着至关重要的作用，信息的交换和共享的范围和有效性是电子政务平台的发展核心的推动力量。以下我们将从电子政务信息交换和信息共享两种方式对电子政务数据交换和共享平台技术架构的进行分析。

由于有数千个点点到点集成在位于全球几十个不同位置的数百个独立系统之间移动，因此要想跟踪每一组数据几乎是不可能的。在这种大规模、不受管理的系统中，不可避免地会发生数据丢失、流程失败以及数据沿袭变得无法跟踪的情况。

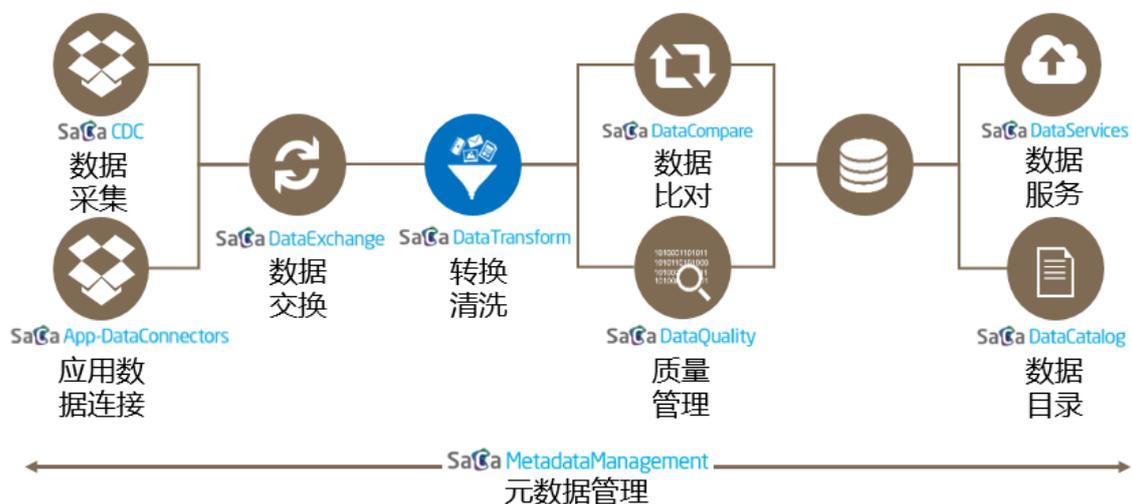
数据交换可管理、监控和跟踪与每个发布和订阅接口关联的事件和元数据。强大的通知和警报系统会将发生的任何问题实时通知操作员，以便数据可以一直移动而永远不会丢失。

SaCa DataExchange，数据交换平台，在 SaCa DI 套件中提供数据交换的能力。



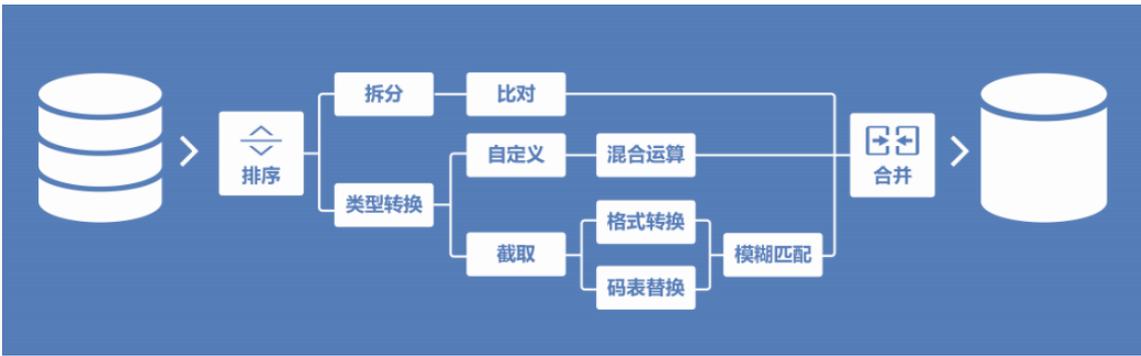
SaCa DataExchange 提供面向复杂网络环境和多样化数据环境的数据交换服务，为企业数据交换提供支撑，可用于快捷打造跨网络、跨部门的数据交换、数据中心建设、数据备份等解决方案。

4. 转换清洗



数据整合的核心内容是从数据源中抽取数据，然后对这些数据进行转化，最终加载到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。

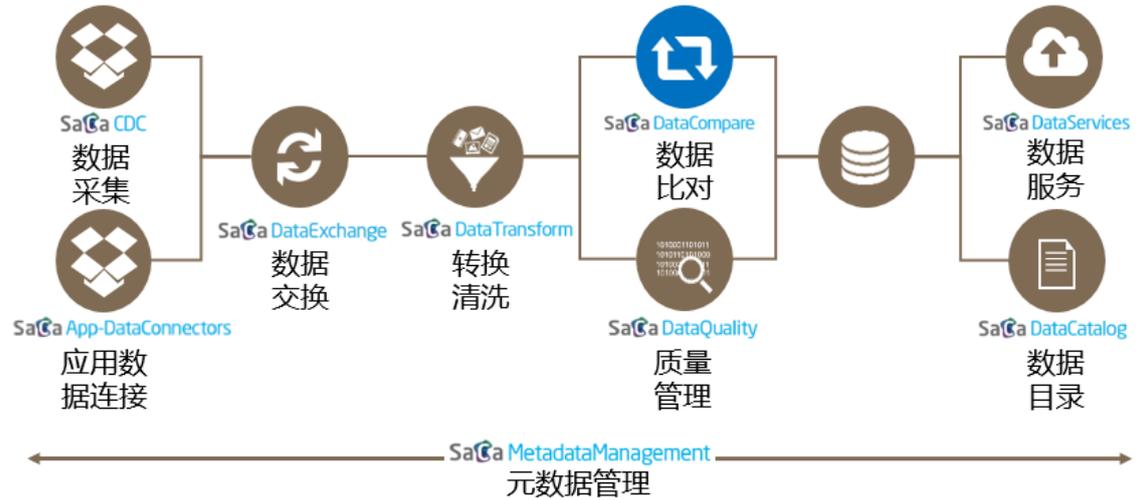
SaCa DataTransform，数据转换与清洗平台，在 SaCa DI 套件中提供 ETL 的能力。



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。

遵从国际化标准以及国家众多行业标准，内置大量数据转换模型，包括几十种通用的转换规则，比如类型转换、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。

5. 数据比对

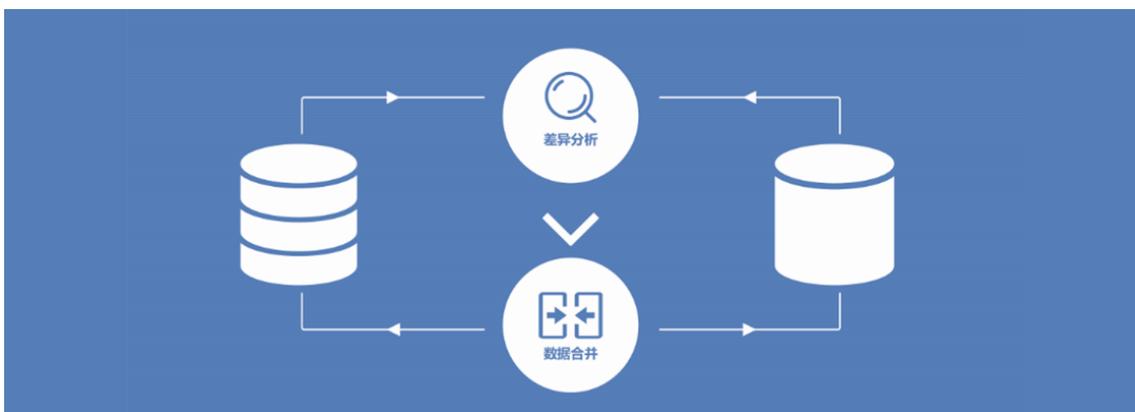


在企业数据中心、灾备中心的建设中，需广泛应用到数据同步复制和数据异步复制技术，以得到共享或灾备数据。不幸的现实是可能出现数据差异而又很难及时发现，这些不一致会导致采用不准确的数据进行数据恢复、基于不准确数据进行企业决策。

当目标数据库中的数据偏离源数据库时就出现了数据差异。数据偏离的程度取决于各种

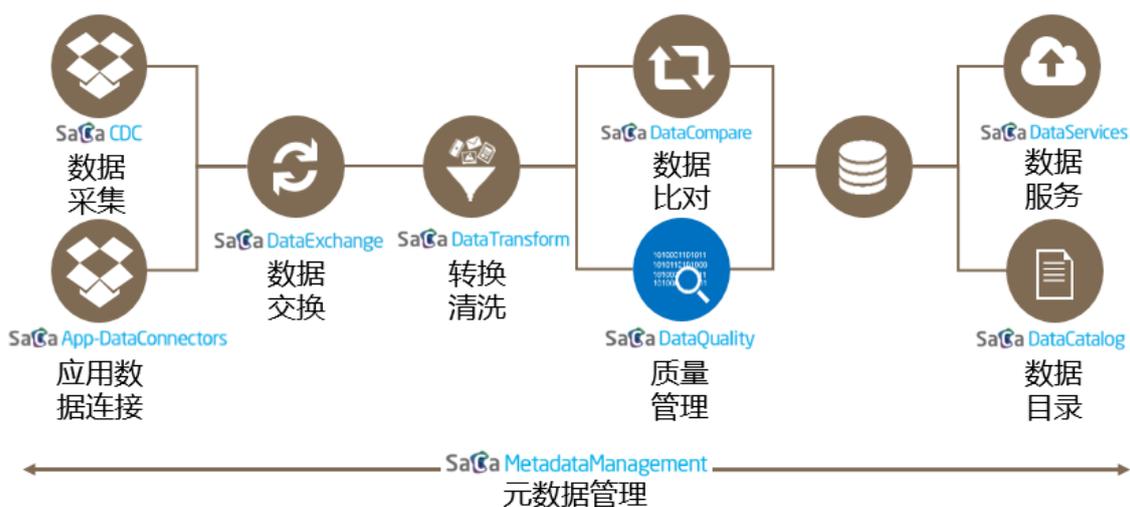
因素，一些可能是有意而为之，一些可能是无意形成的。

SaCa DataCompare，数据比对平台，在 SaCa DI 套件中提供数据差异分析的能力。



在 SaCa DataCompare 的帮助下，企业可以信心十足地在其备份系统、报告数据库、主数据库和其他类型的冗余数据系统中实现数据一致性。并可与 SaCa CDC，SaCa DataTransform，SaCa DataExchange 一起部署，也可以单独部署，能够确保在各数据库之间保持数据一致性。SaCa DataCompare 是一款易于使用、高性能且侵入性极低的产品，可帮助管理整个企业内的数据一致性。该产品可在多种情境下使用，以确保数据一致性，包括比较包含大量数据的数据库或比较数据不断变化的数据库。

6. 数据质量



数据质量是大小企业面临的日益严重问题，较差的数据质量将导致成本增加，缺陷数据

还会妨碍业务决策制定和满足合规性要求而做出的努力。但提高数据质量并不能一劳永逸。数据的问题在于它的质量会随着时间的推移而下降。因此，就像数据采集，转换清洗，交换集成是一个需要持续进行的流程一样，管理数据质量也是如此。数据质量管理涉及的内容不仅仅是查找和修复丢失或者错误的数据库。它还意味着交付值得组织机构信任的数据，也就是说，无论数据的应用程序、使用目的或者来源是什么，始终需要为企业提供全面的、一致的、相关的、适用的数据。

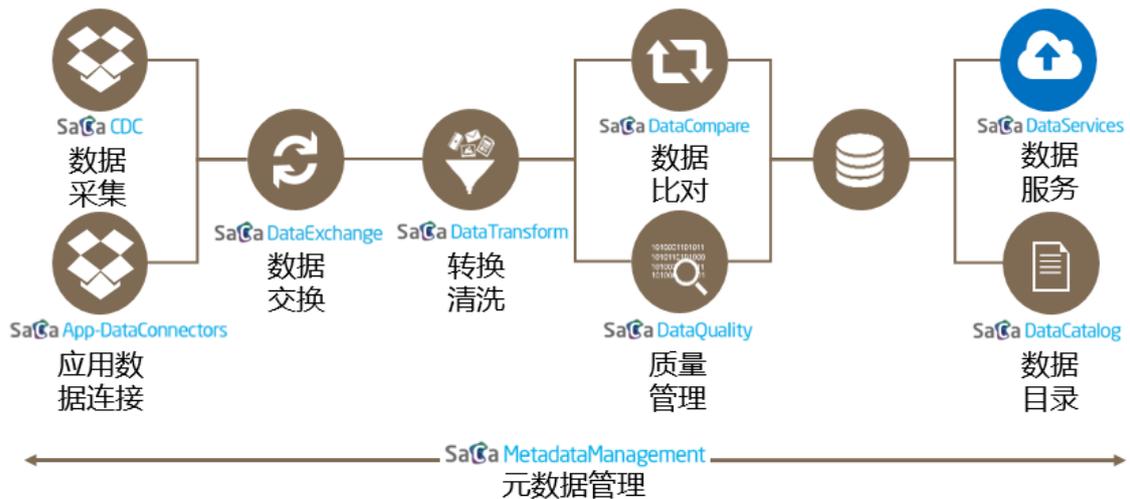
SaCa DataQuality ,数据质量管理平台 ,在 SaCa DI 套件中提供数据质量管理的能力。



SaCa DataQuality 是一款独特的、易于使用的数据库质量软件产品，专为更好地消除 IT 部门和业务部门之间的隔阂而设计，并满足他们能够控制企业范围内数据库质量流程的所有需求，从而降低成本、增加收入和改善企业决策。通过大量内置的数据库质量规则模型，及时发现数据的完整性、一致性、准确性等质量问题，帮助企业改善数据库质量。通过一个统一平台，为所有项目和业务应用程序的利益相关者、项目和数据域（无论在内部，还是在云中），提供权威和值得信赖的数据库质量。它针对您的所有需求，始终保持数据库的可信。

SaCa DataQuality 产品可让业务部门和 IT 部门有效地协作，实现组织机构数据库资产的真实价值。

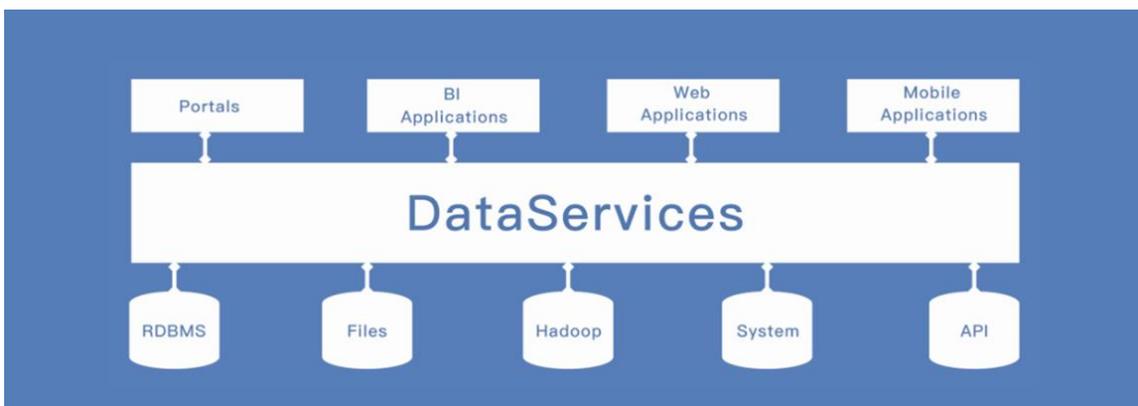
7. 数据服务



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开放和共享。

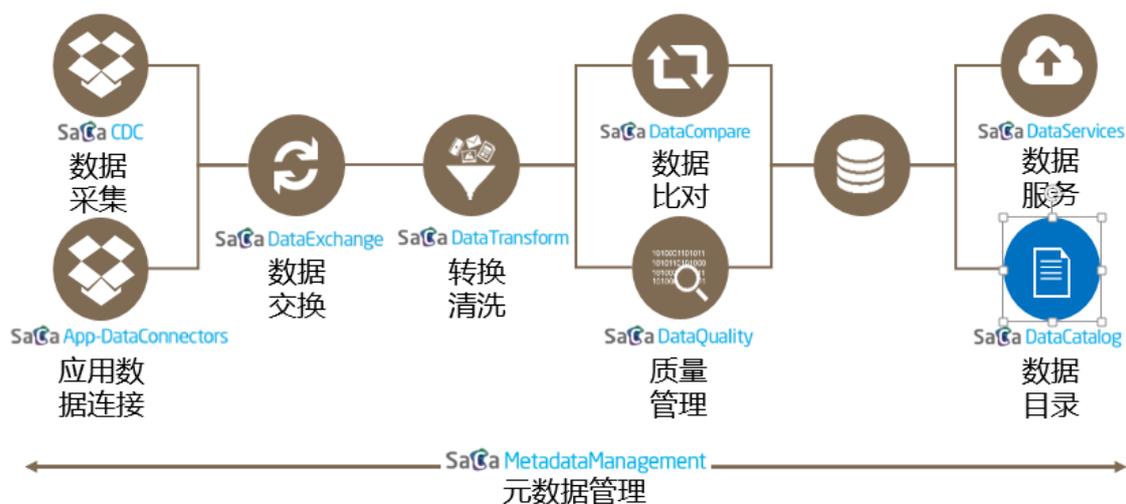
由于组织结构和管理模式等方面的原因，不同职能部门间在过去都是分别独立建立系统，缺乏统一的协调和规划，往往只从具体业务出发，实现手工作业到电子化的转变，很少考虑系统间的信息的共享机制，导致系统数量多，异构化，数据量大、复杂性大。从服务的角度看，就形成服务对象的信息分布在多个业务系统中，难以形成统一的视图。随着管理的变化，很难形成一套整合统一的平台，运转起来困难。同时内部系统目前和外部联系也比较难：重要数据无法共享，数据不是完全共享，特别是一些该共享的重要数据无法共享。另一方面从企业的管理职能看，难以获得管理对象全面的信息，从而影响企业管理决策的效率和准确性。

SaCa DataServices 数据服务平台在 SaCa DI 套件中提供数据服务的能力。



SaCa DataServices 是一款敏捷的数据虚拟化平台产品，通过实时统一的数据访问入口提供数据服务，屏蔽异构数据的复杂性，同时控制数据访问与数据内容的安全性。广泛应用于即席查询、报表的实时数据展现、虚拟数据仓库，Web 方式数据共享等应用场景。

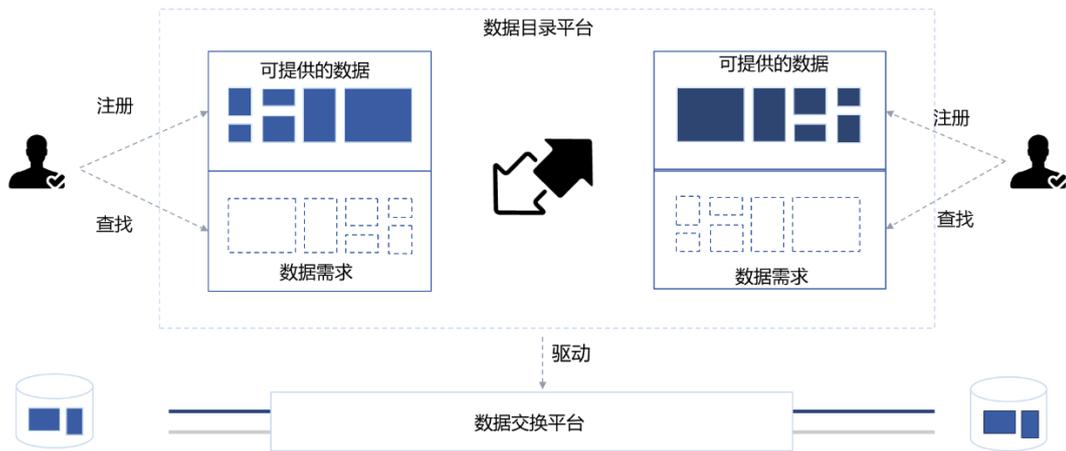
8. 数据目录



信息资源，特别是政务信息资源，是一个国家信息资源的主要组成部分，它是政府部门在履行管理国家行政事务职责的业务过程中和政务信息化过程中产生的、有利用价值的、数字化、网络化的信息内容。有效管理、合理开发和利用政务信息资源对促进政府职能转变、提高行政效率、增强监管和服务能力、推动全社会信息化持续健康发展具有重要意义。

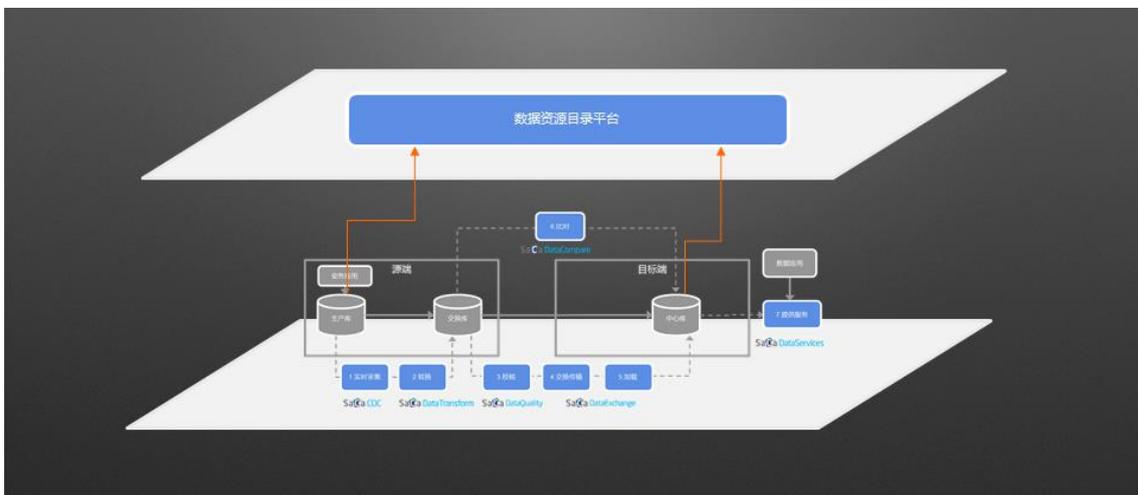
目录体系是开发利用政务信息资源的主要基础设施之一，是政务信息资源共享的重要手段，利用目录技术和元数据技术，以及其它网络技术，在网上构造一个统一的信息资源目录管理系统，使资源的使用者能够在任何时间，任何地点，通过特定的服务接口查询资源目录，使其能快速发现、定位和获取所需信息。

SaCa DataCatalog 数据资源目录平台，在 SaCa DI 套件中提供数据目录的能力。

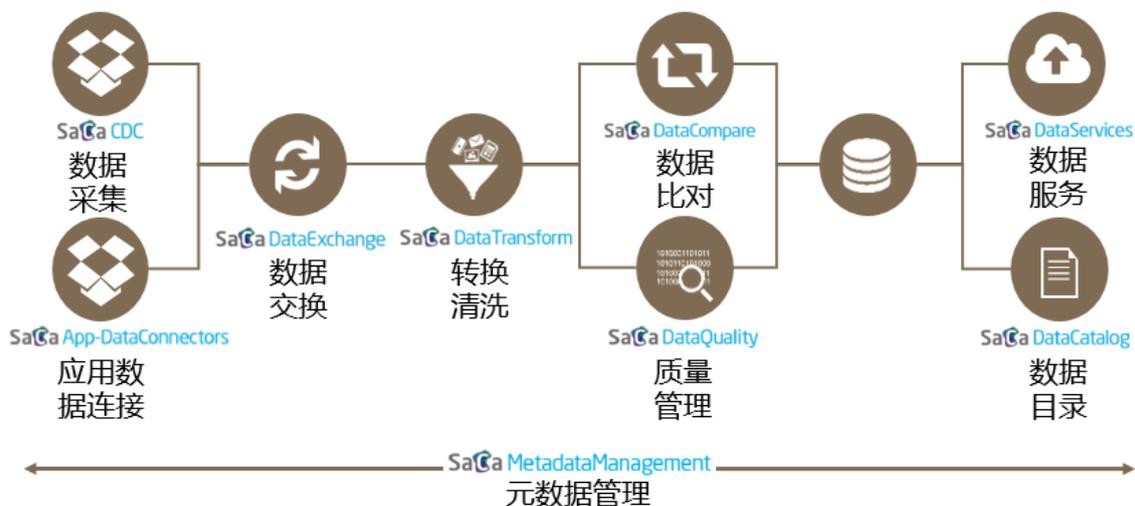


SaCa DataCatalog 数据资源目录平台，面向企业单位及政府部门的数据资源开放共享、数据交换、数据分析等场景，为数据提供方与数据需求方之间共享数据提供平台，提供统一的注册、发现、查询、定位、获取数据资源的入口。组合方案

SaCa DataIntegration 数据集成套件，经过数年多个版本的研发和磨练已经形成了一套系统、成熟的产品体系与组合，将作为原始数据与数据应用的快速通道，帮助用户从原有系统中，包括遗留的数据源，关系型数据库，文本文件系统等，近乎实时的移植、整合、转换清洗、比对、共享各种应用数据。



9. 元数据管理



东软 SaCa MetadataManagement 产品是一款基于 CWM (Common Warehouse Metamodel) 标准规范开发的成熟的企业级元数据管理系统。产品适用于企业复杂的 IT 异构环境和多厂商环境，提供完善的元数据采集、存储、管理和应用功能，支持元数据分析功能，从而帮助企业更好地开展元数据管理的相关活动，可以作为企业信息化的核心基础设施之一。

SaCa MetadataManagement 产品的核心功能包括：

- 自动获取元数据

东软 SaCa MetadataManagement 产品支持多种元数据的自动采集模式，包括模板采集和数据库直连的采集，也提供任务定期调度功能，满足用户周期性采集元数据的需求，保证元数据的可靠性。

其中，采集的数据库类型包括了国内外主流的关系型数据库如 Oracle、MySQL、PostgreSQL、SQLServer、DB2、Sybase、神通、达梦、金仓、GBase8t/8a、虚谷、浪潮、Greenplum 等，以及 HBase、MongoDB 和 HIVE 等。

- 端到端的元数据管理

用户可以依据业务流向查看端到端的元数据信息，支持全生命周期的元数据管理能力，

包括元数据创建、修改、删除等，具备完善的审计能力，可以清晰地查看生命周期内的变更情况。

- 智能语义检索快速定位资产

产品内置检索引擎，支持通过不同维度检索并定位数据资产。额外提供强大的智能语义搜索功能，通过搜索业务术语查找其相关的数据资产，并使用全方位关联分析协助定位资产。

通过高级检索+智能语义检索，彻底解决数据治理中“我关注的数据在哪里”的问题。

- 强大的分析功能

针对任意数据资产，产品提供其全方位关系视图展现，智能分析其直接或间接关联的技术、业务、管理数据资产。借此，元数据使用者可以做到“举一反三”，根据自己的需求逐层分析定位资产，更清晰的了解数据。

尤其针对数据载体，产品具备强大的血缘分析和影响分析等能力，可以满足用户跨 IT 系统和应用的元数据分析需求。产品以业务数据的流向为脉络，可以清晰地展现元数据的流向，以及元数据变更对下游系统和数据的影响。

分析功能在为数据提供可靠的验证手段，提升数据可用性的同时，也为用户规避了元数据变更对下游系统带来的风险，确保了企业整体业务系统的健壮性和可靠性。

四、 总结

SaCa Data Integration 数据集成套件基于一个开放式、与平台无关的体系架构，它可以处理更广泛的数据，为复杂的异构环境量身定做，它可以提供无与伦比的性能，伸缩性和可靠性，适用于企业级的关键任务部署。

五、 东软集团简介

东软创立于 1991 年，是中国第一家上市的软件公司，一直以来致力于以信息
东软内部公开 东软集团股份有限公司

技术的创新，推动社会的发展，创造美好生活。目前，东软在全球拥有近 20000 名员工，在中国建立了覆盖 60 多个城市的研发、销售及服务网络，在美国、日本、欧洲等地设有子公司。

东软连续四次入选普华永道“全球软件百强企业”，在最新一次排名中再度成为入选该榜单唯一的中国软件企业。此外，东软还曾荣获最具全球竞争力中国公司 20 强、中国 50 强全球挑战者、亚洲最受赏识的知识型企业、亚太地区最佳雇主等奖项。

东软以软件技术为核心，面向全球市场提供智慧城市、医疗健康与社会保障、智能汽车互联以及企业互联等领域的创新型解决方案、产品与服务。

在智慧城市领域，东软聚焦政府、金融、能源、电信、制造等二十多个行业，提供 210 多个业务方向、750 余种产品与解决方案，在众多行业拥有领先的市场份额。东软以“惠民、优政、兴业”为目标，构建绿色、开放、共享的新型智慧城市。

在医疗健康与社会保障领域，东软构建了以人为核心的可持续发展的大健康生态系统，推动信息技术与医疗的融合，创新医疗的新模式。面向居民及患者、医药服务支付方、政府监管方、医疗服务提供方以及生态链各方，提供线上、线下相结合的全生命周期的产品与解决方案。东软的软件支持了中国 7 亿人群的社会保险服务，医疗设备销往全球 110 多个国家和地区的 9000 多家医疗机构，健康管理业务服务居民超过 3000 万人。

在智能汽车互联领域，东软拥有 28 年服务全球客户的经验，以软件为驱动，推动汽车产业的开放与融合，专注电动汽车、智能网联、自动驾驶与共享汽车领域，致力于打造安全、舒适、智能、绿色的汽车生态系统和移动出行服务。东软在该领域的全球研发网络拥有 4000 名工程师，软件服务于全球 top30 汽车品牌中的 85%。

在企业互联领域，东软基于移动互联网、人工智能、云计算、大数据等新技术，提供企业互联产品以及专业的云与数据服务，帮助企业实现业务的创新与转型。

东软一直秉承“超越技术”的经营思想和品牌承诺，关注人，关注社会发展，致力于成为受社会、客户、股东和员工尊敬的公司。

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

