

解决的主要技术问题

UDIS 产品所解决的主要问题也是目前数据应用和 ETL 产品所面临的问题。

目前基于数据的应用市场已经比较明确，但大部分实施的数据应用项目存在以下问题：

■ 缺少完整的软件架构支撑，存在“孤岛”

主要表现在以下两个方面：

一是“存储孤岛”普遍存在着“应用开发与数据存储单一对应”的问题，往往是一个新的应用开发，便相应地增加一个信息存储产品，不同应用产生的信息存储在不同的模块中，由于分别存储，在一定程度上影响了数据共享。

二是“应用孤岛”由于在设计时没有采用开放的体系架构，各系统体系结构往往是“蜂窝煤”式的，系统之间互相隔离，其应用不能被其他系统所集成。

■ 缺少相应的工具来统一数据标准和规范数据

缺少支持数据统一的工具，不方便数据格式的标准化和规范化。

在业务上主要表现在以下两个方面：

一是数据定义和口径不一致：不同用户对同一数据名称或统计口径有不同理解，对不同口径的产生原因又缺少明确的解释和说明；同一统计口径在系统中有不同的计算方式。

二是指标结构单一：缺乏分析模型和专业化的深入分析来验证指标的合理性；不能根据不同类型的特征，按行业和种类采用针对性的预警指标。

究其根源主要的还是缺少技术手段或工具来帮助实现数据模型的规范和标准。

■ 缺少软件服务和组件的管理

数据应用必然需要和产生大量的软件服务模块和服务组件，如何更好的管理和重用这些服务和组件将变得非常重要，也大量的节省 BI 的开发和维护成本。

目前数据应用，业务信息没有集中管理，多头应用、多头发布，重用性差，主要表现在以下两个方面：

一是报表多头发布，缺少集中管理和发布渠道，造成了信息发布的不规范和低效率

二是没有集中的、专用的数据应用服务，重用性差，各系统仍需承担大量的重复的查询、统计等应用，无法“瘦身”，影响了系统运行效率，也增大了运营成本；

■ 用户定制不方便

数据应用项目是和用户的需求紧密相关的，用户的需求和场景是经常变化的，因此满足用户的个性化定制将变的非常重要。目前数据应用项目在个性户定制方面主要表现在以下两个方面的问题：

一是系统后端部分缺少统一的架构和统一开放的应用模型定义，有的虽然采用了模型，但模型不完整，效果不好。这样用户在使用时，不能根据其需求动态的调整后端的业务规则和运行环境，不利于用户的使用。

二是前端部分组件化差，缺少前端模型和模板，有的虽然采用了模板，但缺少模型描述，重用性不好。这样用户在使用时，不能根据其需求动态的调整前端显示组件的组合，不利于用户的扩展使用。

■ 数据应用适时性差

数据的时效性将和数据的应用效果紧密相关的，如何提供快速、高效的、适时的处理结

果是用户非常关注的，也往往决定数据应用项目的成败。而目前的很多数据应用在时效性方面主要表现在以下两个方面问题：

一是数据来源的适时性抽取差，不能监控源数据的变化，不能做变化数据的抽取，不能做适时抽取。

二是数据抽取和数据处理慢，影响了数据应用的效果，由于不能对变化的数据准确把握，因此处理细分的颗粒度比较粗，就增加了数据后续处理的工作量，延长了处理的时间，使本应该日出的结果，变成了周出，周出的结果，变成了月出，极大地影响了应用效果。

■ 元数据管理差

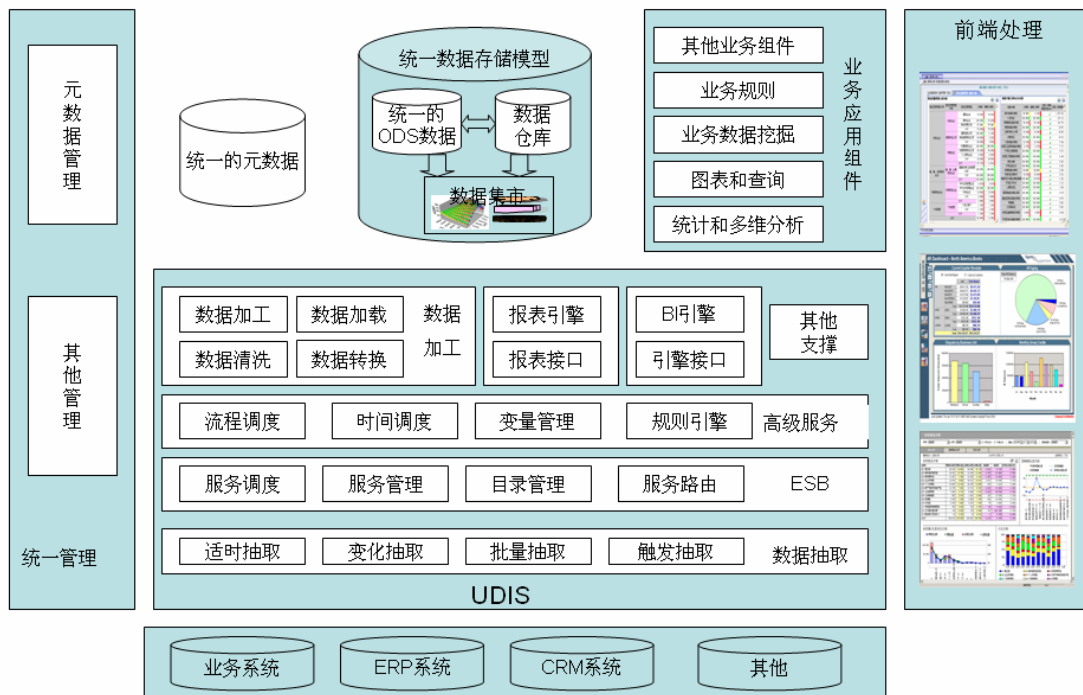
元数据是描述数据的数据，涉及数据处理、规则、表等相关信息。元数据是很有价值的信息，应该很好地描述、处理和重用。数据应用项目会产生大量的元数据，元数据的管理将变得非常重要。目前很多数据应用在元数据管理方面主要存在以下问题：

一是缺少统一管理元数据，目前分散的元数据不利于统一的管理、维护和二次开发。

二是缺少合理分类元数据，目前不合理的分类让元数据的管理比较困难。

三是缺少可视化管理元数据，目前元数据的管理需要比较专业的知识，增加了元数据管理的和二次开发的难度。

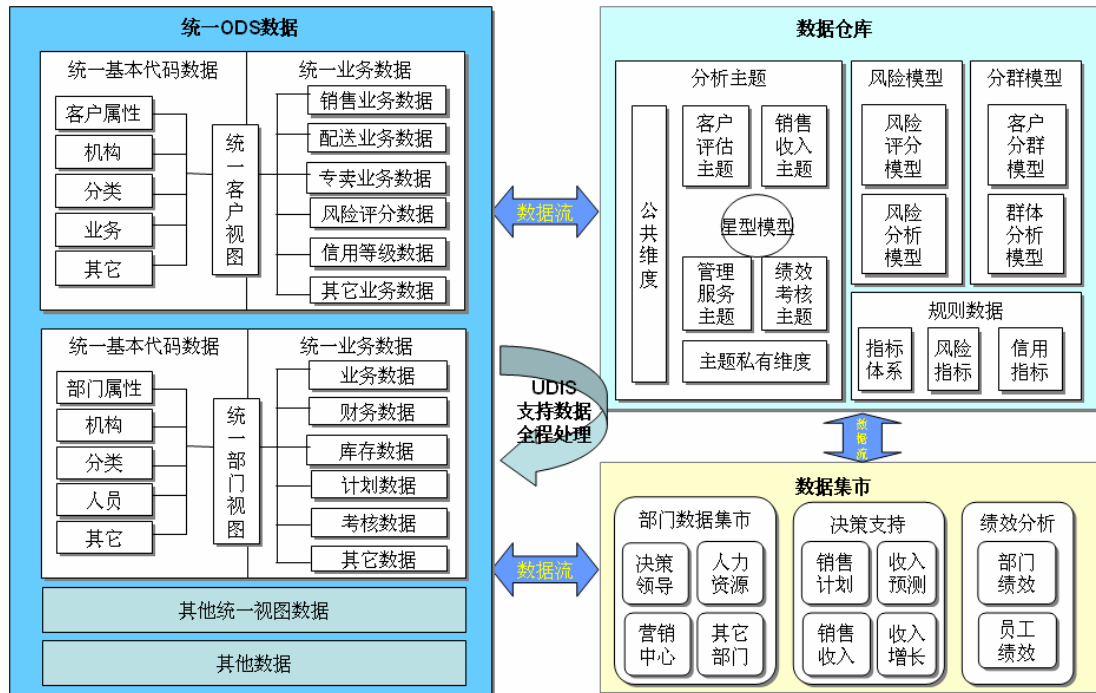
为了很好的解决以上问题，我们提供如下解决方案



该方案主要是基于统一的架构模型，该架构主要由数据源部分、UDIS 部分、统一数据存储模型部分、统一元数据部分、前端业务应用组件部分组成。该方案主要是围绕统一的数据存储模型而服务的，UDIS 是为统一数据存储模型提供和加工数据的数据推送引擎。

统一数据存储模型提供了基于业务的统一数据标准和规范化数据，包括统一的 ODS 数据模型、数据仓库模型、数据集市等。其中，ODS：存放近期的一些细节数据，用于支持数

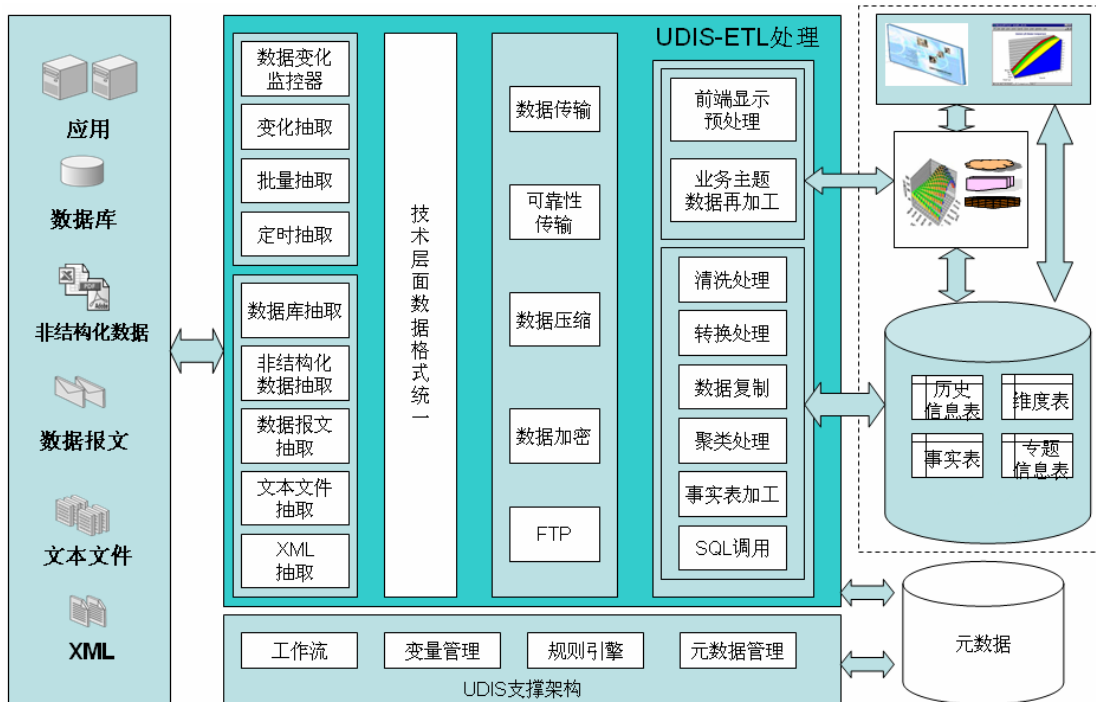
据的明细查询和报表；数据仓库：存放历史的分析数据信息，用于大量的历史分析和数据挖掘；数据集市：存放某一业务领域的的数据，用于支持该领域的数据分析和挖掘等。



UDIS 支持统一数据存储模型

数据处理路径可以有如上图所示的几种路径。数据的加工和处理主要由统一数据集成服务产品 UDIS 来实现。

UDIS 是支持“全程数据推送”的数据集成产品，其 ETL 部分 UDIS-ETL 主要实现数据的抽取、传输、清洗、转换、装载等处理服务。如下所示：



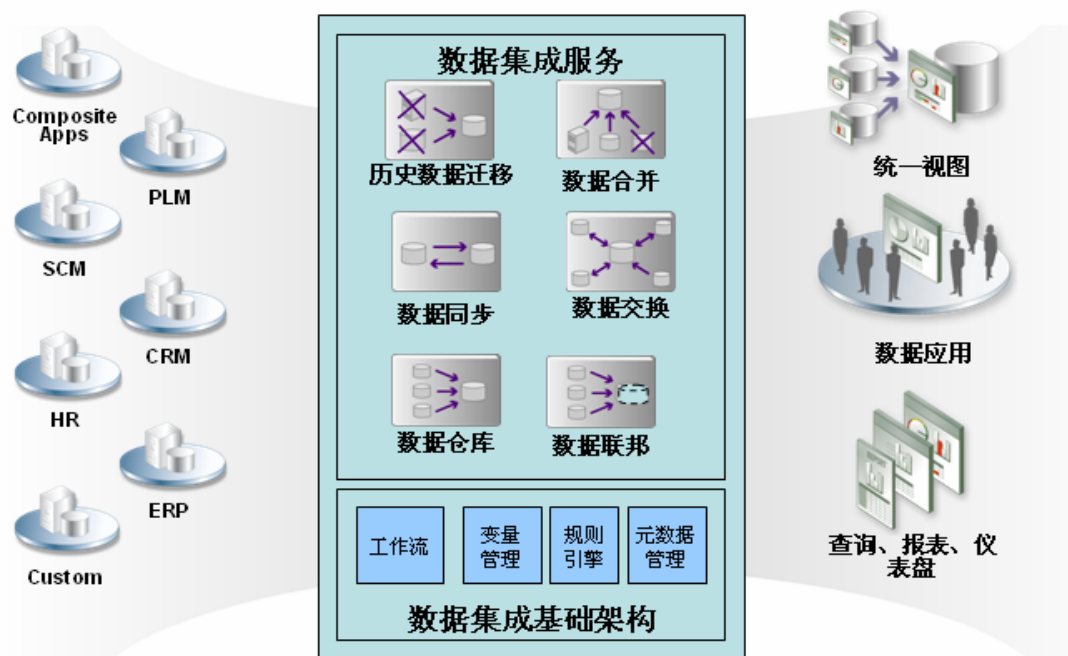
UDIS-ETL 支持“全程数据推送”

在这里，数据的抽取可以基于批量的抽取、变化的抽取、适时的抽取方式抽取数据，抽取后的数据如果位于广域网或分布式的网络环境下可以做分布式的数据传输，做可靠的数据传输处理，然后将数据的复制加载到 ODS 库中，然后调 predefined 的转换处理、SQL 调用等作相应的加工；再调用对数据仓库的实事表加工、业务主体的再加工等处理给数据仓库中的实事表或专题信息表加载数据；仓库完成后，还可以调用相应的数据集市加工服务，对数据集市加载数据，也可以调用前端显示的预处理，给显示部分加载合适的结果集。当然，一些 BI 项目可能仅用到其中的几步操作。

UDIS 在这个解决方案中主要起到：

- 帮助数据模型的规范化、标准化，给统一的数据模型提供经过加工处理的规范化、标准化的数据，并给数据模型的规范化、标准化提供帮助工具。
- 实现全程数据的推送处理，支持从数据的抽取、传送、清洗、转换、实事表加工、专题数据的再加工等一条龙服务，所有过程自动完成，并将处理后的数据推送到结果集中。
- 提供规则引擎，满足基于规则的数据抽取、专题数据的加工等需要。
- 提供元数据的集中管理，将数据加工过程中生成的元数据合理分类、并将数据加工处理模型化，生成预定义模型，方便使用模型的使用，更好的重用业务人员的知识积累。

主要的应用范围



UDIS 应用场景图

UDIS 是数据集成服务产品，用于从企业已有的业务系统中集成数据，并给分析型应用或操作型应用提供数据，满足企业数据分析和数据应用的需要。UDIS 主要包括数据集成服

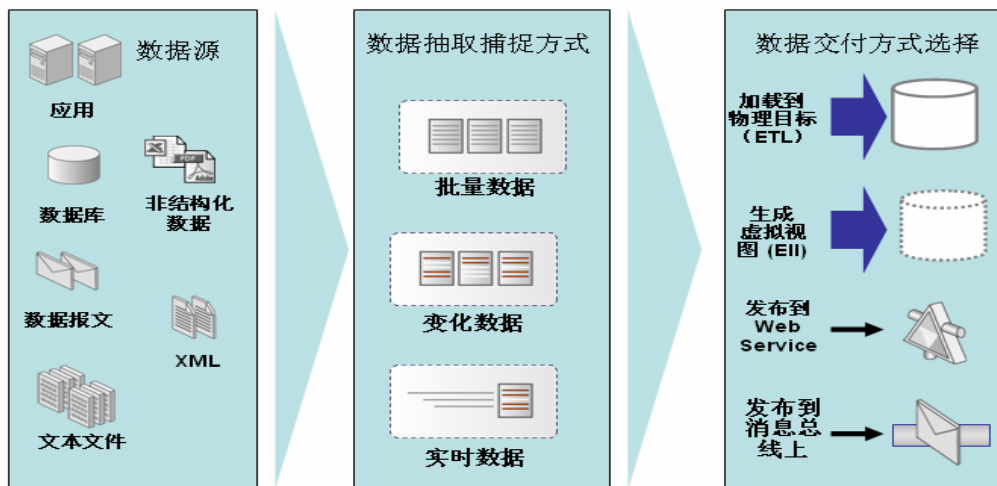
务和数据集成基础架构。

数据集成服务主要包括历史数据迁移、数据同步、数据合并、数据交换、数据仓库、数据联邦等服务。

- **历史数据迁移**：实现历史遗留数据的重用，将历史数据迁移到新的目标数据库中。
- **数据同步**：实现分布的数据库中数据的上传、下载，保证分布数据库中的数据的一致性。
- **数据合并**：实现不同应用的数据库中的个性化数据到统一结构的数据库中的数据的合并，包括清洗、转换等操作。
- **数据交换**：实现分布的数据库和集中的数据库之间的交互。包括数据集中、数据分发、格式转化等。
- **数据仓库**：实现将分散的数据集中到统一的数据仓库中，并建立统一的数据模型来存储。包括：数据的上传、转换、将处理后的结果装载到事实表中，结合维度表形成数据立方。供 OLAP 分析、报表、预测等 BI 应用使用
- **数据联邦**：保存数据源的映射关系，数据仍在原系统中存储，主要满足数据的实时处理、统一结构视图等需要。

数据集成基础架构是各种数据集成服务的运行环境，各种数据集成服务作为插件“插入”到基础架构中，由架构实现对集成服务的查找激活、输入/输出参数生成、生命周期的管理、运行的监控、意外的处理等。该架构内置了 workflow、变量管理器、规则引擎、元数据管理等部件。

- **工作流**：实现数据集成处理的流程自动化，基于流程自动化可以实现数据应用项目中“数据全程推送”。工作流引擎可以自动调用每个处理节点的数据服务，同时通过调用变量管理器给数据集成服务提供输入参数；工作流支持数据加工并行处理、串行处理、混合处理。
- **变量管理器**：实现数据集成服务的变量的描述和实例的生成，主要满足工作流处理过程中各处理节点变量参数输入的需要，满足规则引擎中动态表达式的变量实例的生成需要。
- **规则引擎**：通过提供规则服务将业务处理中所需的规则独立出来处理，满足业务规则预定义和运行时规则更改的需要。可以满足流程服务和业务对规则的使用需求，业务人员按需求定义业务规则，将规则及相关设定录入规则模型库中。
- **元数据管理**：实现对数据集成服务、流程处理、源数据库、目标数据库等模型的描述、配置、编辑等。给运行引擎提供运行元数据模型。各元数据是以 XML 格式描述，方便提供第三方使用，同时也方便引入第三方的元数据。



UDIS 集成方式选择

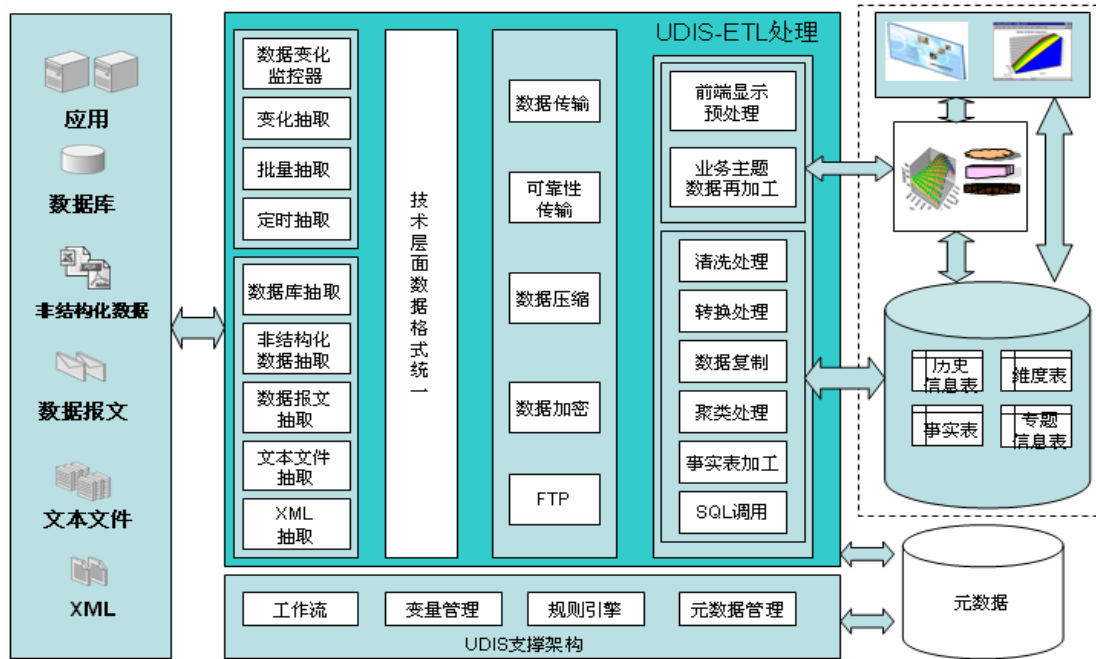
UDIS 为了方便数据应用的实际需要，提供灵活的数据抽取捕捉方式、数据交付方式，方便数据应用项目根据实际情况选择不同的数据集成方案。其中，数据抽取捕捉方式支持：

- 批量数据抽取：定期批量的抽取数据源的数据
- 变化数据抽取：根据数据源的变换，抽取变化的数据（包括：新增、修改、删除等）
- 实时数据抽取：实时捕捉数据源的数据，并抽取。

数据交付方式支持：

- 加载到物理目标的 ETL 方式：将抽取过滤的数据经过清洗、转换处理后加载到目标库中，这里可以进行事实表和立方体的处理，也可以仅作转换处理到目标表，也可以不作转换直接到目标表等。
- 生成虚拟视图：做数据联邦处理，不做数据的集中，数据分散存储在各自的业务系统中，目的是提供统一视图供其他系统使用数据。
- 发布到 Web Service：将集成的数据发布到 Web Service 供其他系统使用。
- 发布到消息总线上：将集成的数据发布到消息总线上供其他系统使用。

其中，ETL 处理是 UDIS 很重要的功能之一，其功能图如下所示。



UDIS-ETL 处理功能图

UDIS-ETL 处理主要包括数据抽取层、数据的传输层、数据的加工层。数据抽取层主要实现数据源的批量抽取、变化抽取、适时抽取，支持数据库、非结构化数据、数据报文、文本文件、XML 等多种数据源，经过数据抽取后，将各种数据来源的数据在技术层面上实现格式统一；数据的传输层实现对于分布式的部署系统的数据传输，可以根据网络的情况选择不同的传输方式，对于网络条件比较差的情况，提供可靠性传输机制，支持数据的压缩、加密、FTP 等处理；数据的加工层实现传输后的数据的加工，主要包括数据的清洗处理、转换处理、数据的复制、数据的聚类处理、事实表加工和 SQL 调用等，同时还支持数据的再加工，可以为数据的前端显示等操作作预处理。