

ESENSOFT



小亿直播



小亿直播

数据治理之元数据管理

汇报人：王超

时间：2020.06.28



目录

CONTENTS

I

元数据管理背景

II

元数据管理方案

III

亿信元数据管理工具

IV

亿信元数据成功案例



I 元数据管理背景

A hand holding a tablet displaying data charts. The charts include a bar chart with red and blue bars and a line chart with multiple colored lines. The background is a dark blue gradient with a large, faint watermark of the ESENSOFT logo and the text '元数据管理背景'.

斯诺登事件

美国国家安全局搜集了与电话呼叫相关的元数据

亚历山大图书馆的卷录

公元前245年，世界上第一个元数据系统

剑桥分析事件

Facebook积累的大量的用户网络社交元数据，进行网络分析，精准投放广告方式，“助攻”特朗普竞选

卡片式目录

每条记录成为可以独立操控的对象





一间摆满书的屋子能被称为图书馆吗？
要在图书馆中找到某本书，不能只是在图书馆中走来走去，单靠运气去找。

- 图书馆无法管理的自己图书，很难统计馆内有多少图书、每类图书多少
- 图书馆无法根据大众读者喜好摆放某类图书的位置
- 读者无法找到自己想读的图书
- 读者费时费力地找到了图书，但内容与描述不符
- 读者精疲力尽地找到了图书，但无法快速定位到某些章节
- 读者心平气和地找到了图书，但内容是错误的
- 读者心满意足地找到了图书，但内容是下册的，又必须从上册读起
- 读者喜出望外地找到了图书，但内容是用外文写的，用外文作的注解
- 读者欲哭无泪地找到了图书，但图书馆要下班关门了
-读者崩溃了.....



找不到数据、数据找到了，没有上下文无法理解数据

理解了数据但是因为数据格式问题无法使用

内容有误导致结果错误

数据查询性能低

数据加工好已经错过时效

如果没有元数据管理，数据无法被有效地组织起来、被准确地理解、被合理地使用和产出预期的结果，那么数据价值无法发挥出来；如果没有元数据，那么数据的内容和真实性就难以估量，继而可能造成数据价值和可用性的降低。元数据是发挥数据价值的前提，是数据治理的基石。

元数据：描述数据的数据

元数据分类：业务元数据、技术元数据、管理元数据

业务元数据：面向业务分析人员。业务元数据是描述数据的业务含义、业务规则等。通过明确业务元数据让人们更容易理解和使用业务元数据，元数据消除了数据二义性，让人们的数据有一致的认证，避免“各说自话”，进而为数据分析和应用提供支撑。常见的业务元数据包括：业务定义、业务术语、业务规则、业务指标等

技术元数据：技术元数据是对数据的结构化，方便计算机或数据库之间对数据进行识别、存储、传输和交换。技术元数据可以服务于开发人员，让开发人员对数据的存储、结构更明确，从而为应用的开发和系统的集成奠定基础。技术元数据也可服务于业务人员，通过元数据理清数据关系，让业务人员能够更快速的找到想要的的数据，进而对数据的来源去向进行分析，支持数据血缘追溯和影响分析。常见的技术元数据包括：存储位置、数据模型、数据库表、字段长度、字段类型、ETL脚本、SQL脚本、接口程序、数据关系等

管理元数据：面向IT系统运维技术人员。管理元数据描述了数据的管理属性，包括管理部门、管理责任人等，通过明确管理属性，有利于数据管理责任到部门和个人，是数据安全管理的基石。常见的管理元数据包括：数据所有者、数据质量定责、数据安全等级等。

元数据类型	元数据描述	元数据实例
业务定义	数据的含义	客户的完整名称, 并具有法律效力
业务规则	数据录入规则	企业的营业执照、组织机构代码证书, 统一社会信用代码证书等具有法律效力的证明文件中的中文名称全称
	识别规则	企业的组织机构代码或者统一社会信用代码或者统一纳税号必须完全匹配, 则认为是同一客户。
	质量规则	客户名称为非空, 并且与营业执照的中文名称一致
存储位置	数据的存储什么地方	ERP系统
数据库表	存储数据的库表名称和路径	ERP/Customers
字段类型	数据的技术类型	字符型
字段长度	数据存储的最大长度	[200]
更新频率	数据的更新频率	每年更新一次
管理部门	数据责任部门	客户管理部
管理责任人	数据责任部门	客户管理部业务员

“

元数据绝对会告诉你关于别人生活的一切。如果你有足够的元数据，你就不需要数据内容。

”

斯图尔特·贝克尔

美国国家安全局前总法律顾问



元数据管理方案



元模型设计

- 基于国际通用CWM模板
- 元模型客户化设计

元数据采集整合

- 配置采集源
- 设置采集任务
- 补充管理属性及依赖关系等

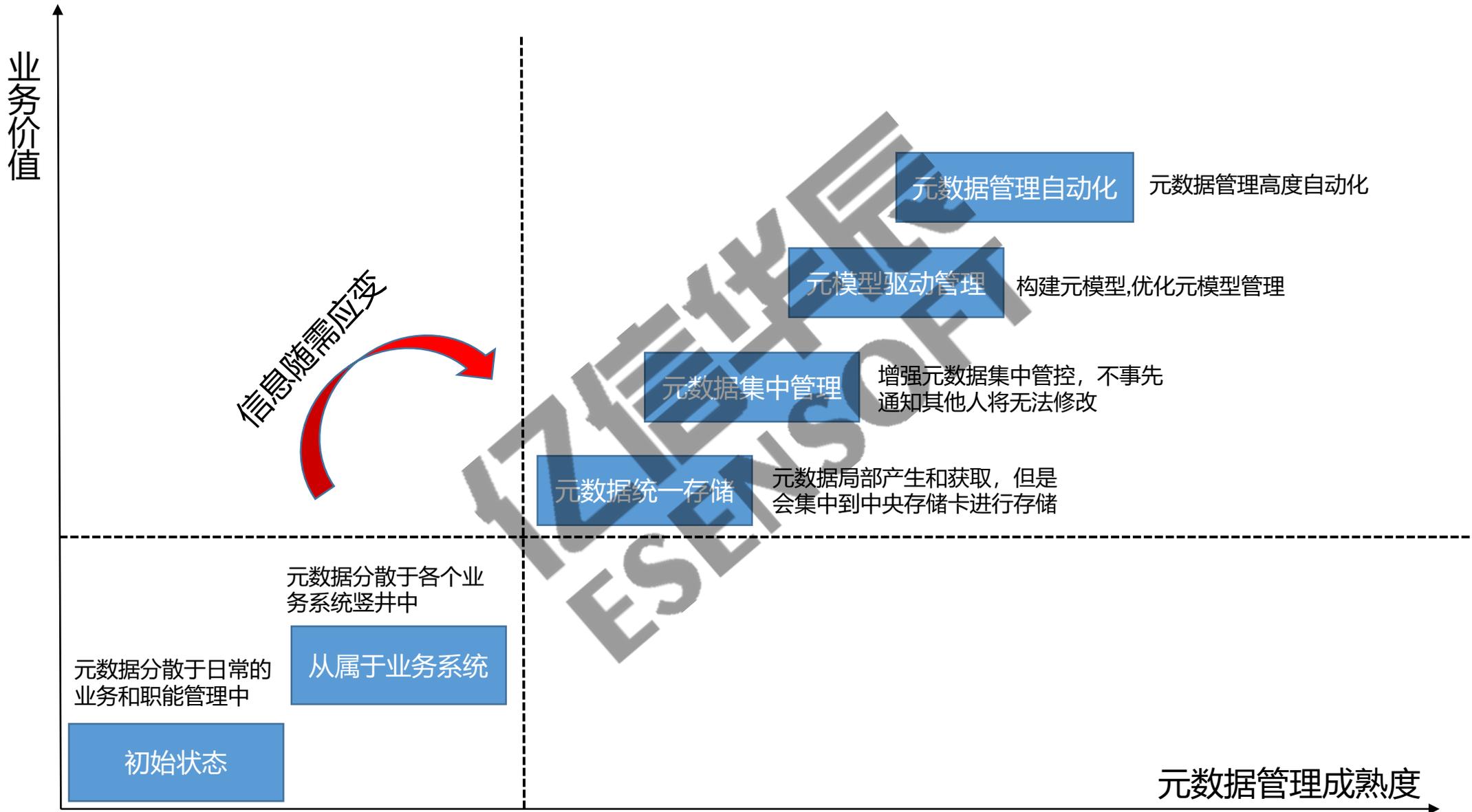
元数据管理

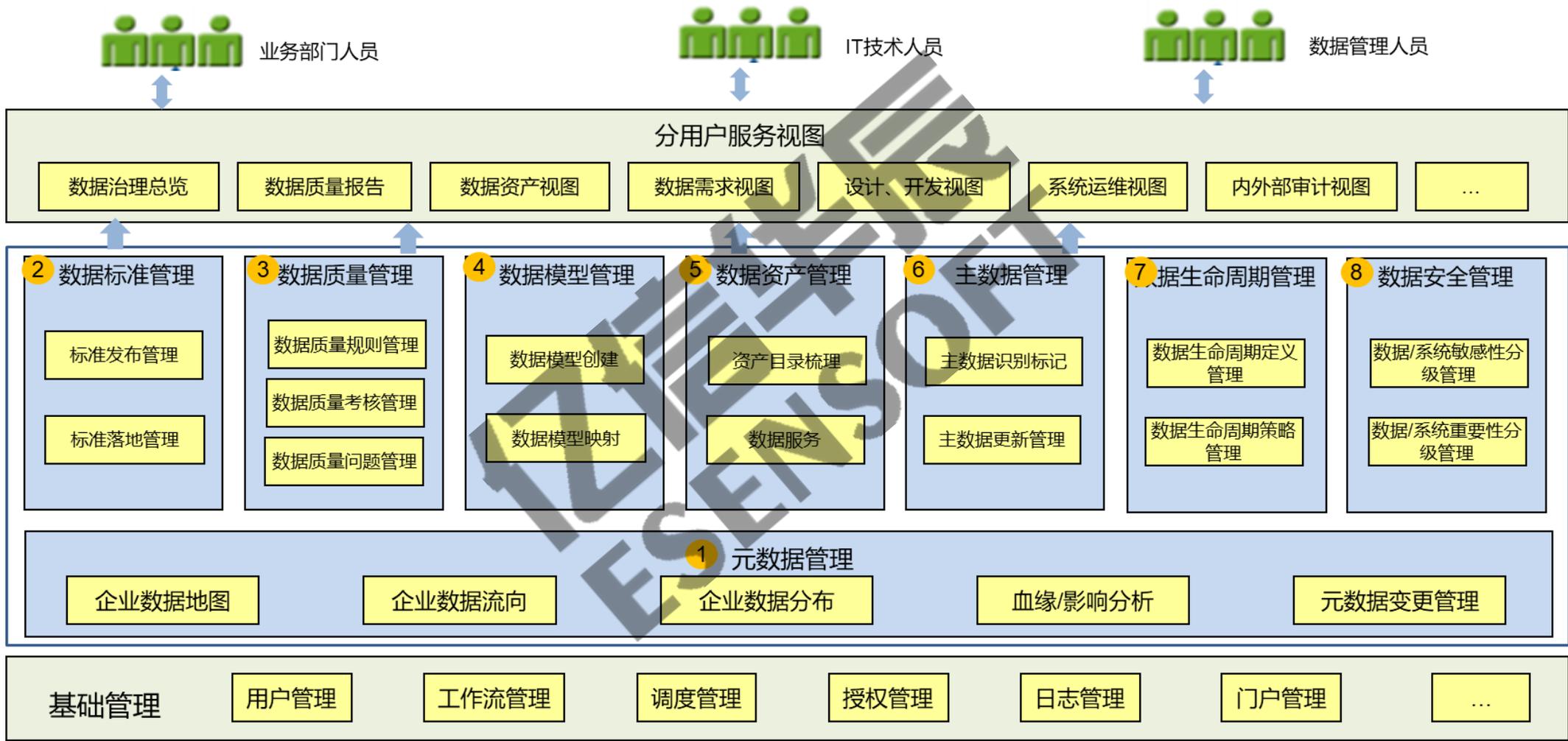
- 定期对比检查一致性
- 定版元数据发布

元数据服务

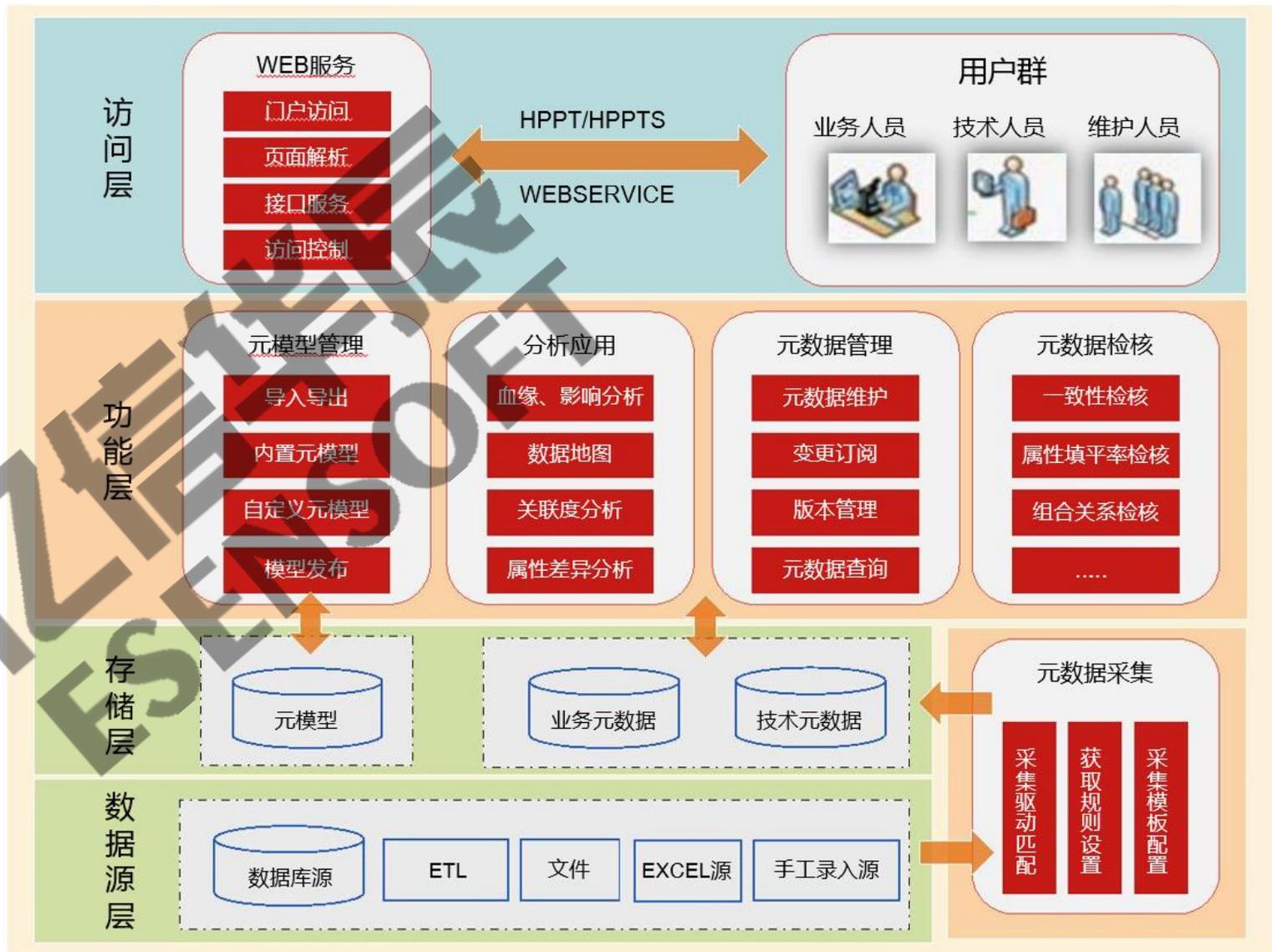
- 元数据查询 - 我是什么、如何定义, 分布在哪
- 血缘分析 - 我是谁加工出来的
- 影响分析 - 我支持了谁的加工
- 数据资产服务

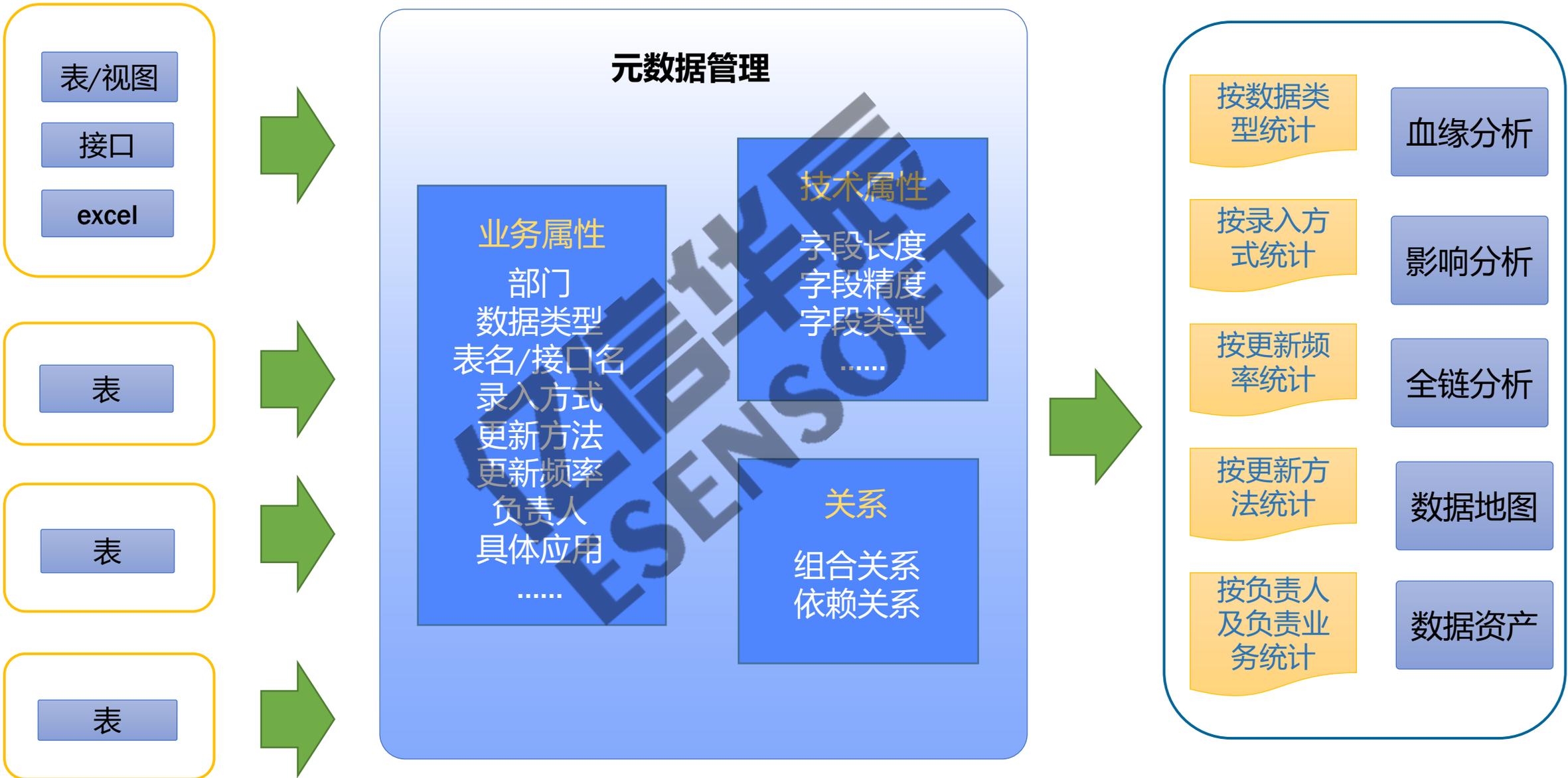
理清数据的来龙去脉





元数据管理平台从应用层面，可以分为：元数据采集存储，元数据管理服务、元数据分析服务以及元数据访问服务。





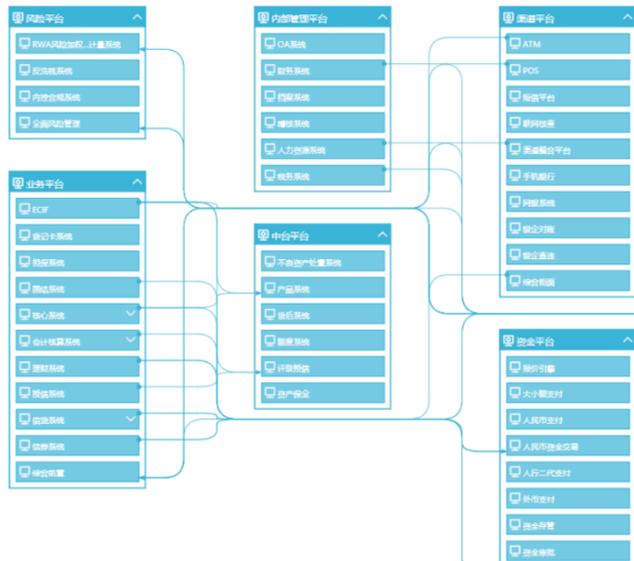
一图在手，天下我有

通过元数据以企业全局视角对企业各业务域的数据资产进行盘点，实现企业数据资源的统一梳理和盘查，有助于发现分布在不同系统、位置的数据，让隐匿的数据显性化。数据地图包括了数据资源的基本信息，存储位置信息、数据结构信息、各数据之间关系信息，数据和人之间的关系信息，数据使用情况信息等，使数据资源信息详细、统一、透明，降低“找数据”的沟通成本，为数据的使用和大数据挖掘提供支撑。

追根溯源，发现数据问题本质

企业在做数据分析的时候，数据分析结果不正确，原因可能是数据分析过程出现数据问题，也可能是数据源本身就有问题，还可能是数据在加工处理过程中出现了数据问题……。通过元数据血缘分析，能够快速定位数据来源和加工处理过程，能够帮助数据分析人员快速定位数据问题。另外，通过元数据血缘关系分析，可以理解不同数据指标间的关系，分析产生指标的数据源头波动情况带来的影响。

最新元数据 定制元数据 重新加载 放大 缩小 重新布局 导出 保存 打印





元数据管理工具

亿信仕展
ESENSOFT

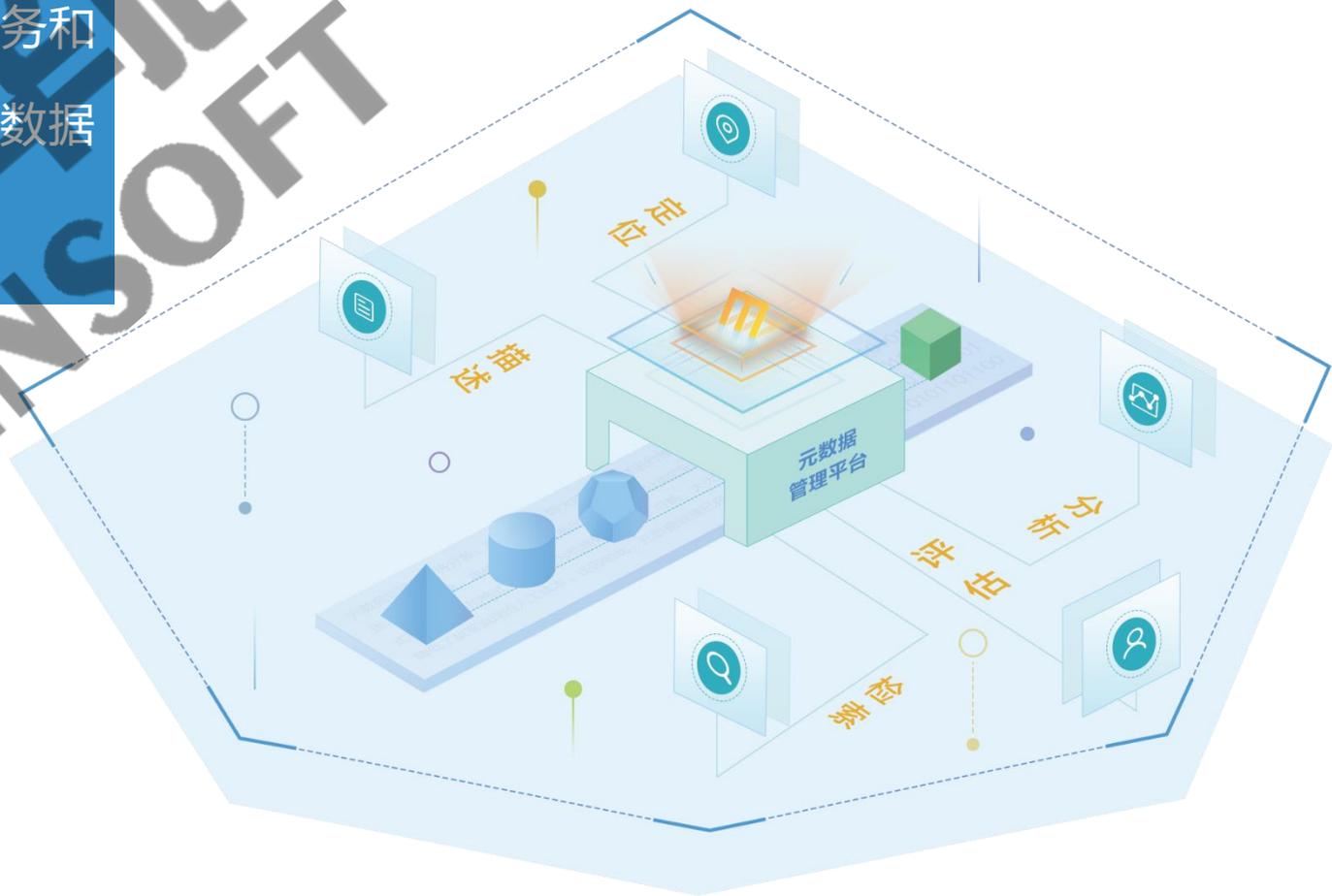


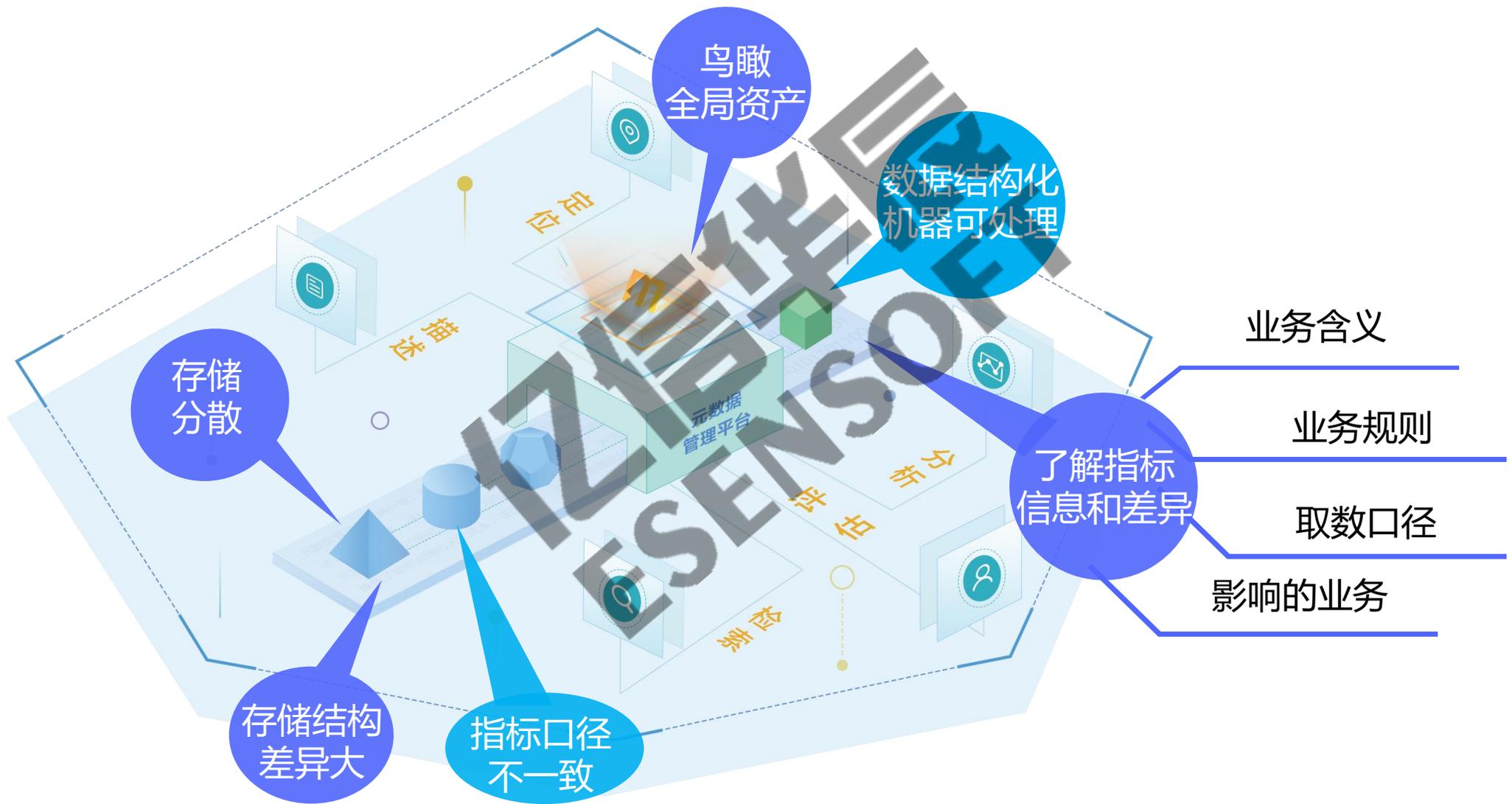
睿治数据治理平台是我公司**完全自主研发的、开创性的、一站式**综合数据治理整体解决方案。睿治是**全国领先，实现了数据治理场景全覆盖的突破性产品**，九大核心模块：**元数据**、数据标准、数据质量、主数据、数据资产、数据安全、数据交换、数据处理、数据生命周期等，**以创新的方式保证了企业的业务数据在采集、汇总、转换、存储、应用整个过程中的完整性、准确性、一致性和时效性，全面为客户量身打造符合自身特征的数据治理体系。**

元数据管理为亿信睿治数据治理平台九大模块之一

亿信元数据管理平台定位

亿信元数据管理平台致力于处理各种技术、业务和操作类元数据，帮助各行各业用户获得更好的数据洞察力，进而挖掘出隐藏在资源中的价值。





先进

基于Web客户端架构

强大的多类型源适配

提供端到端的元数据管理统一视图

可扩展

支持CWM标准的基础上

提供元模型扩展机制

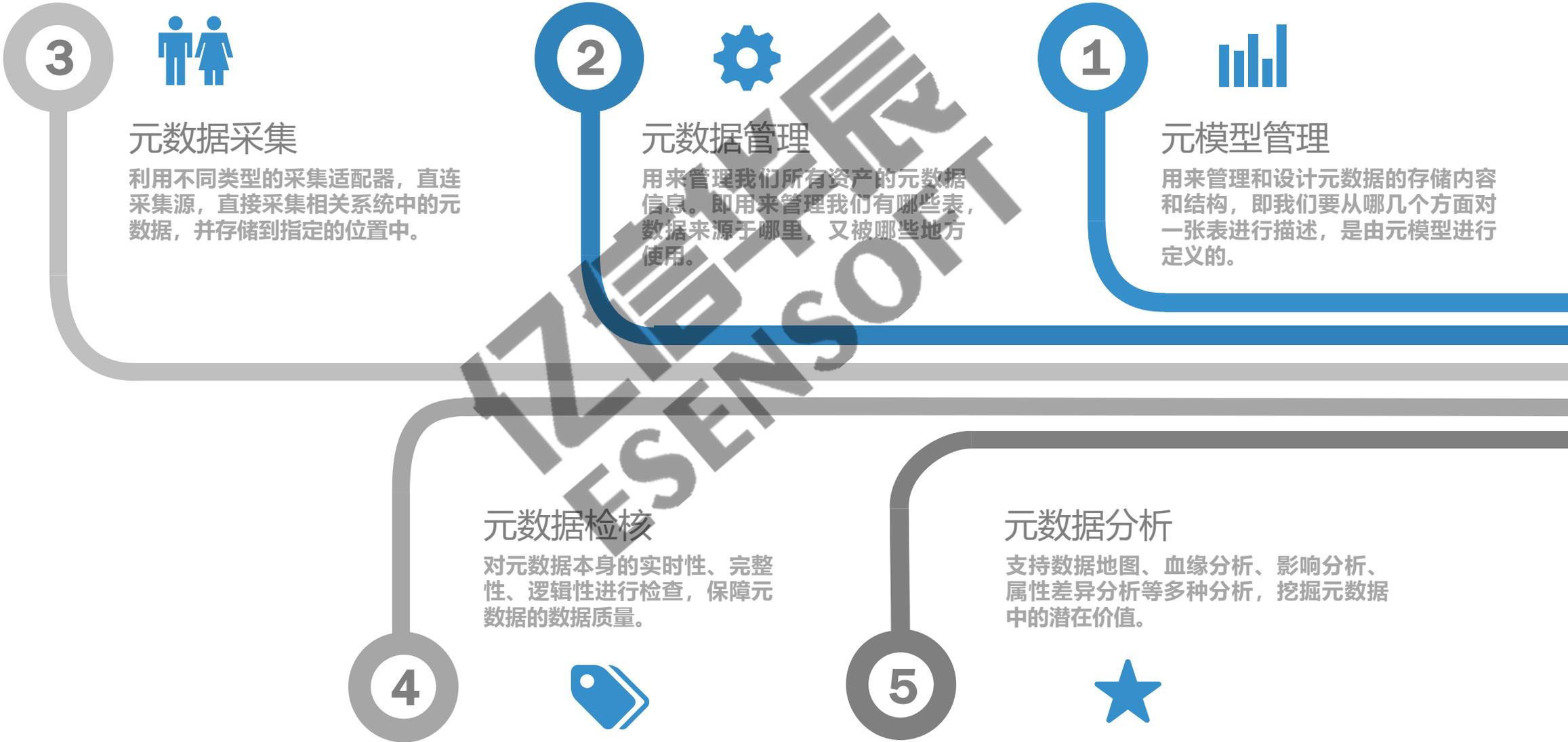
同时做到元数据接口开放，易集成

易用

类互联网式搜索，优秀的可视化技术

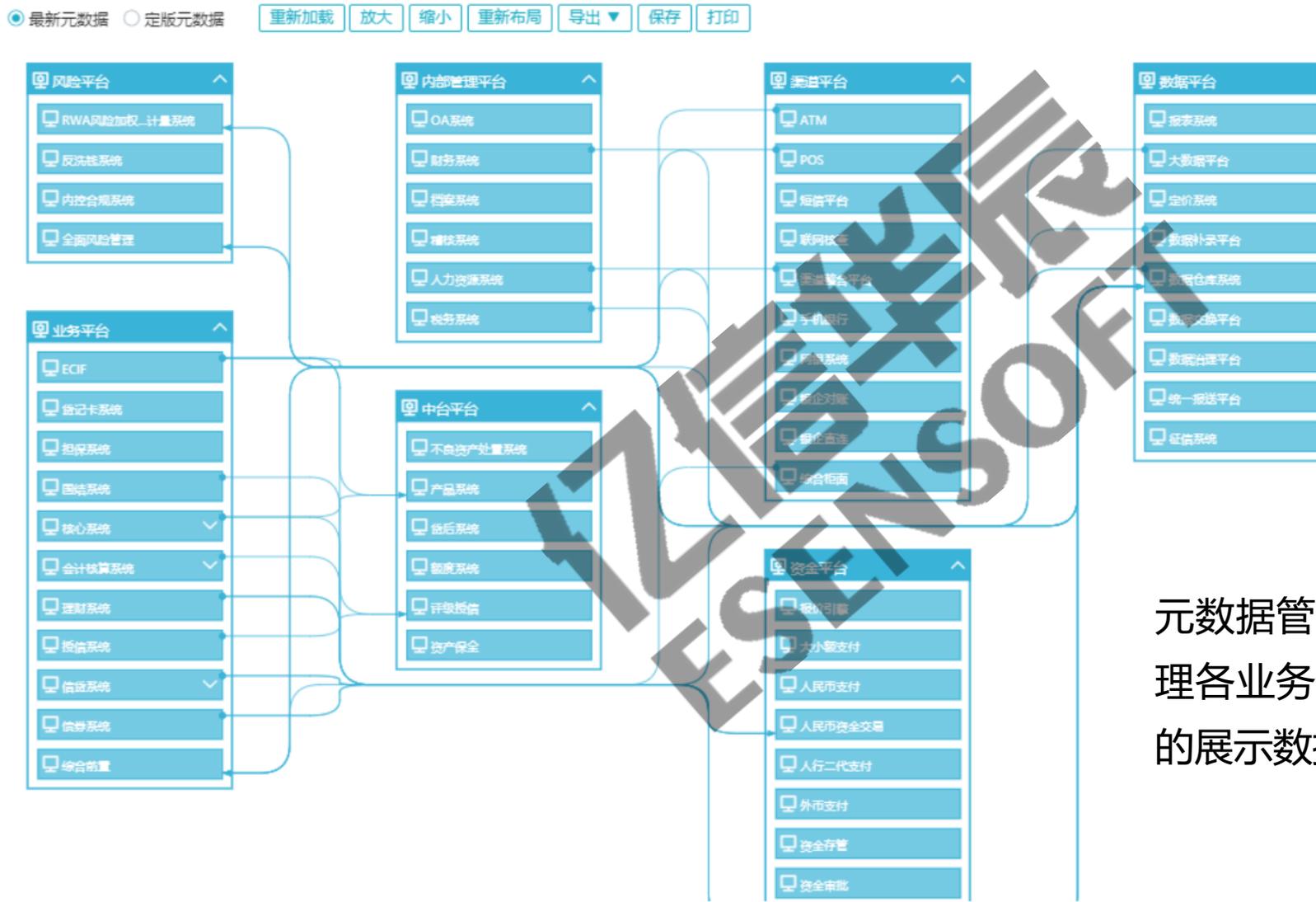
提供元数据全生命周期可视化管理机制

亿信元数据管理平台5大功能模块



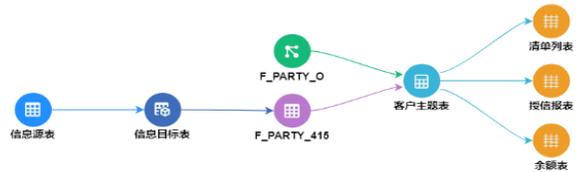


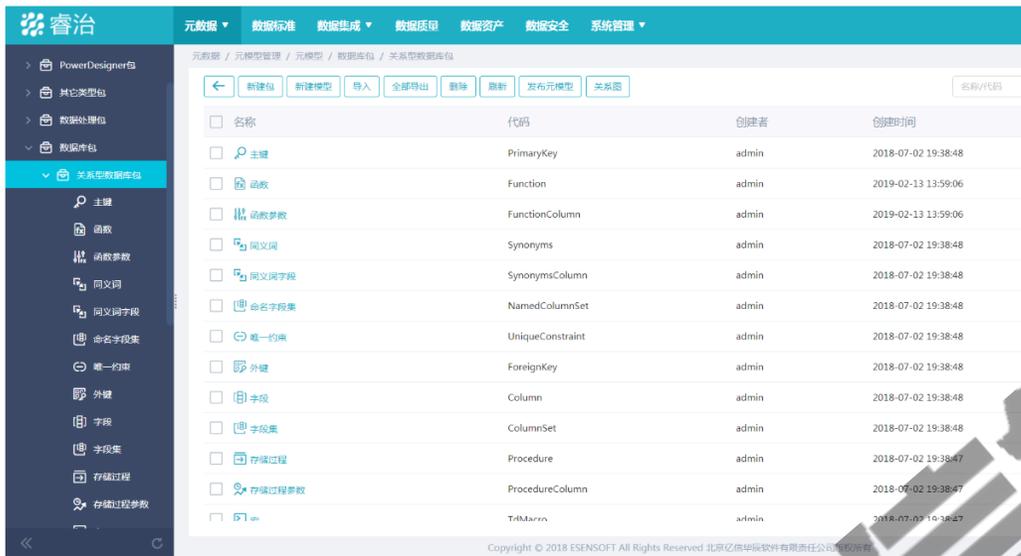
元数据管理平台内置多种采集适配器，支持多种存储格式的元数据自动获取，且适配器采用可扩展设计，最大限度满足用户不同的元数据自动化采集场景。如：数据库、报表工具、ETL工具、文件系统等，同时为无法完成自动获取的元数据，提供了可自定义的元数据采集模版完成元数据的批量导入



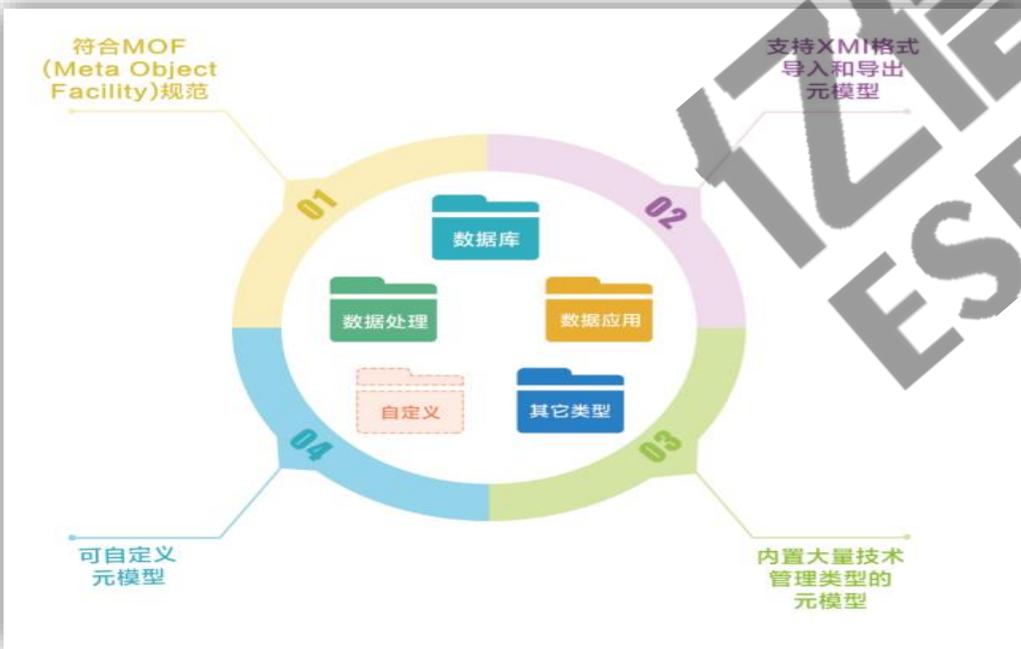
元数据管理平台提供数据地图功能，全面梳理各业务系统的分布情况及关联关系，清晰的展示数据全貌，形成企业数据全景图

元数据管理平台提供了丰富的元数据分析功能，包括血缘分析、影响分析、全链分析、关联度分析、属性值差异分析等，分析出元数据的来龙去脉，快速识别元数据的价值，掌握元数据变更可能造成的影响，以便更有效的评估变化带来的风险，从而帮助用户高效准确的对数据资产进行清理、维护与使用

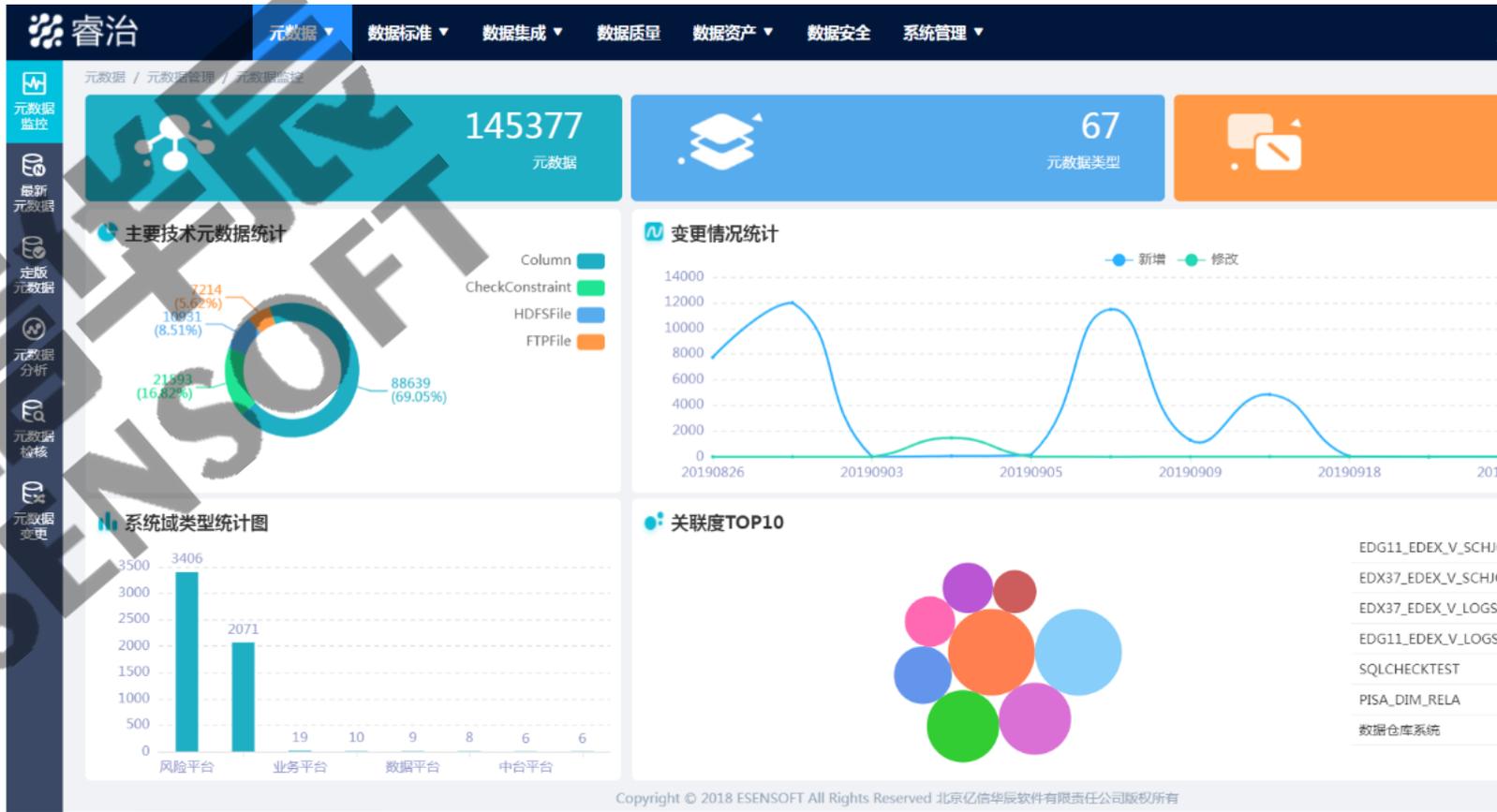




- 元模型的结构遵循了UML规范，按照包类的树形层级结构管理。元模型在支持CWM（公共仓库元模型）国际标准的同时，提供了一套便捷的自定义管理接口功能，支持根据用户管理需要，进行自定义元模型以及元模型之间关系的扩展。
- 支持XMI格式导入和导出元模型
- 内置大量技术、业务、管理类型的元模型
- 可自定义元模型



元数据管理平台提供了完善的元数据监控功能帮助用户一目了然的了解元数据的使用情况和分布情况，以统计图的方式来展现元数据数量，分布情况，近期变更情况，关联度情况等，可视化效果好并且支持按照用户需求来自定义监控界面



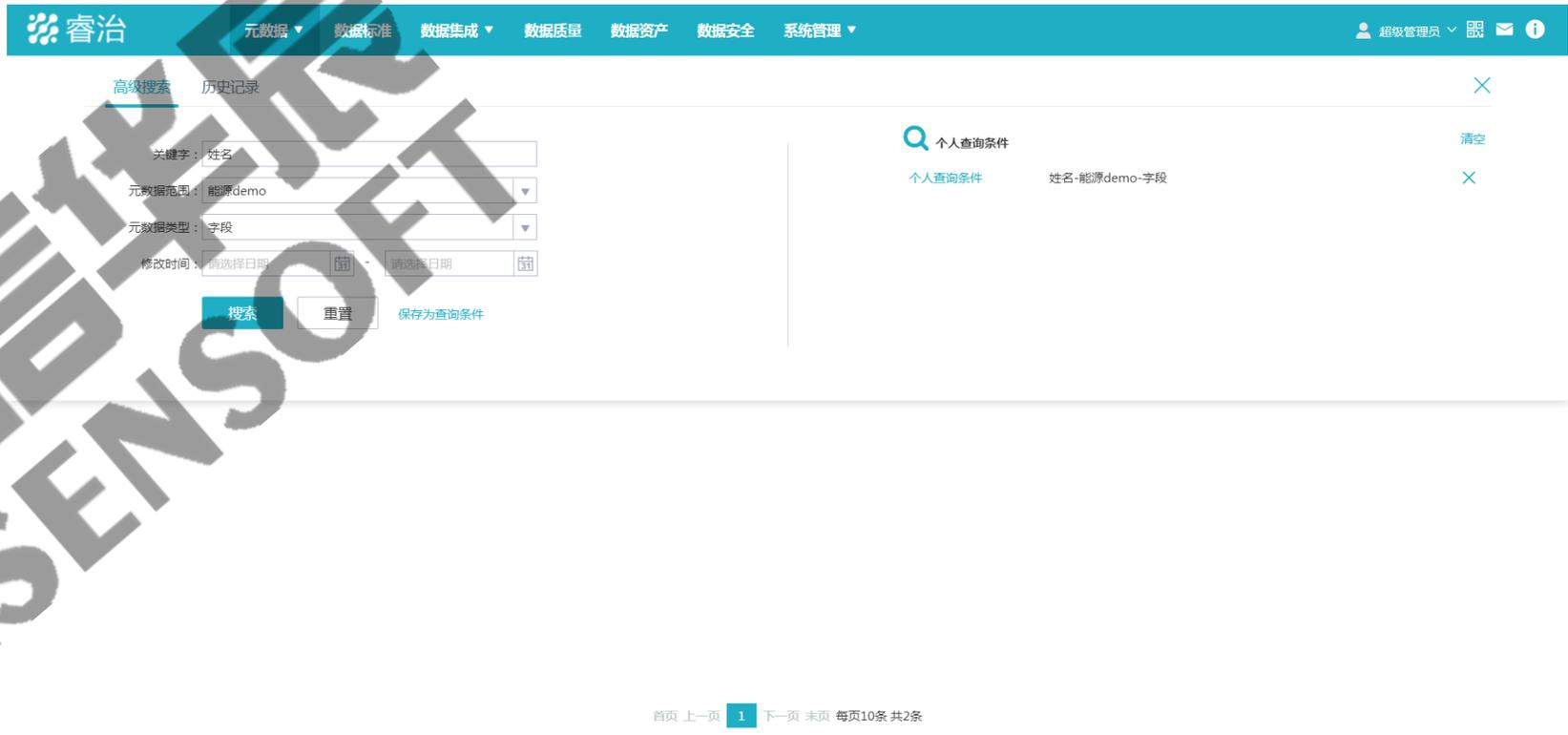
元数据版本比较-列表

基准版本: v1 对比版本: v2 对比

元数据类型	v1	v2	新增	删除
视图	4	4	0	0
检查限制	312	312	0	0
库	1	1	0	0
索引	104	104	0	0
目录	1	1	0	0
主键	64	64	0	0
表	67	68	1	0
触发器	2	2	0	0
字段	1030	1029	0	1
唯一约束	3	3	0	0

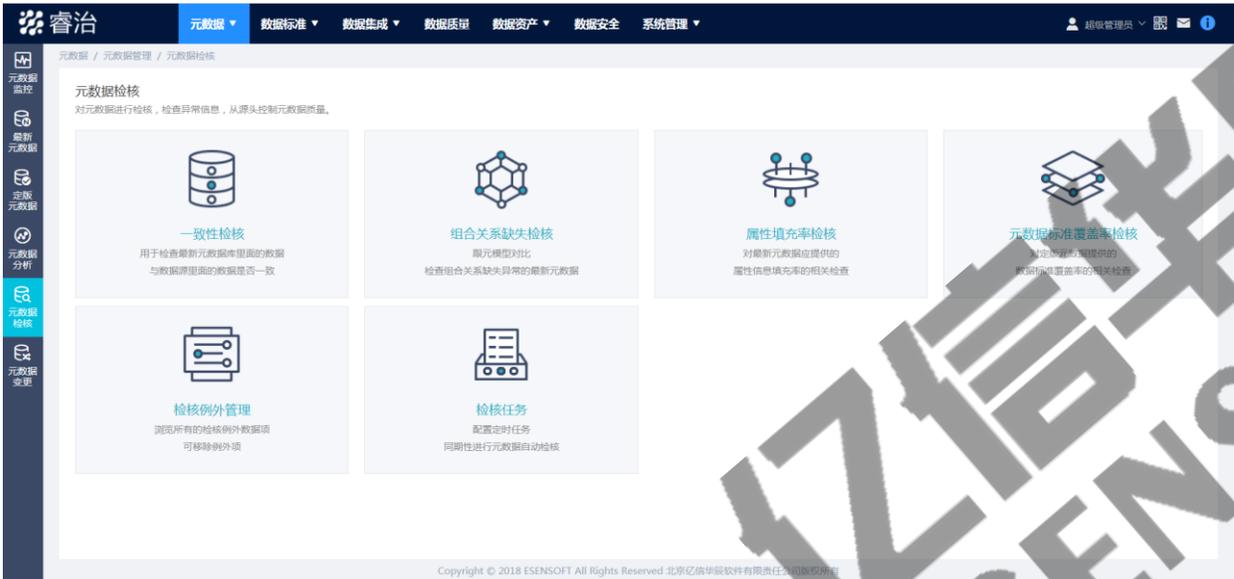
元数据管理平台提供各类元数据管理，包括：业务元数据、技术元数据和管理元数据，支持元数据的基本信息、属性、依赖关系、组合关系的增删改查操作。最新元数据和定版元数据隔离，在最新元数据中的改动不影响定版元数据的正常使用，同时每次发布都有版本留痕，支持各版本的对比分析

元数据管理平台提供了类互联网方式的元数据检索功能，采用的是多级筛选，逐级过滤，由模糊到精确的过程。除了根据简单的名称、标题等方式的模糊匹配外，根据元数据本身的特征，定义了一批可供筛选的条件，让元数据的检索像上购物网站购物一样，非常的方便、快捷，满足用户对元数据各种查询场景的要求



自动检核

人工检核



入库审核
从源头控制质量

元数据管理平台提供元数据质量检核功能，包括一致性检核、属性填充率检核、覆盖率检核和组合关系检核，是保障元数据质量的重要手段之一。



元数据成功案例





该行数据服务类项目主要分为以下两大类：

一是，**报表应用类项目建设**，该类应用旨在实现银监会、人民银行、外管局等外部监管机构数据报送以及内部数据分析。主要包括2009年综合报表平台建设、2013年数据平台项目，2015年领导驾驶舱项目、2016年基础设施信贷监测、2017年综合报表平台升级改造等项目群。

二是，**数据平台类项目建设**，该类项目旨在实现全行数据的集中，建立全行一致数据视图，提升数据质量。主要包括2013年数据平台一期建设，2014年数据平台二期整合层建设，2015年数据平台优化完善、2016年数据平台运维等项目群。

在进行综合报表平台建设过程中，引进和开发了多种数据管理工具，包括：

- 1、**数据标准制定**：针对监管报送和行内管理要求，对行内主要业务指标进行梳理，形成基础和指标数据标准，为后续数据平台建设奠定基础。
- 2、**元数据管理系统**：通过元数据自身属性及其血缘分析、影响分析，对数据的存储、加工和应用过程进行全程跟踪和监控。
- 3、**数据调度平台**：实现对ETL作业依赖关系的统一管理和自动化调度
- 4、**数据质量检查系统**：对数据仓库中的数据质量进行检查，并根据所属系统、主管部门、业务种类等维度生成数据质量检查报告。
- 5、**数据生命周期管理系统**：利用数据库分区技术，实现对数据仓库在线数据的生命周期进行管理。
- 6、**数据切分、传输平台、数据加载**：实现对总行报表数据按省分行级别的切分、下发以及省分行的数据加载，为省行特色报表的提供了数据。

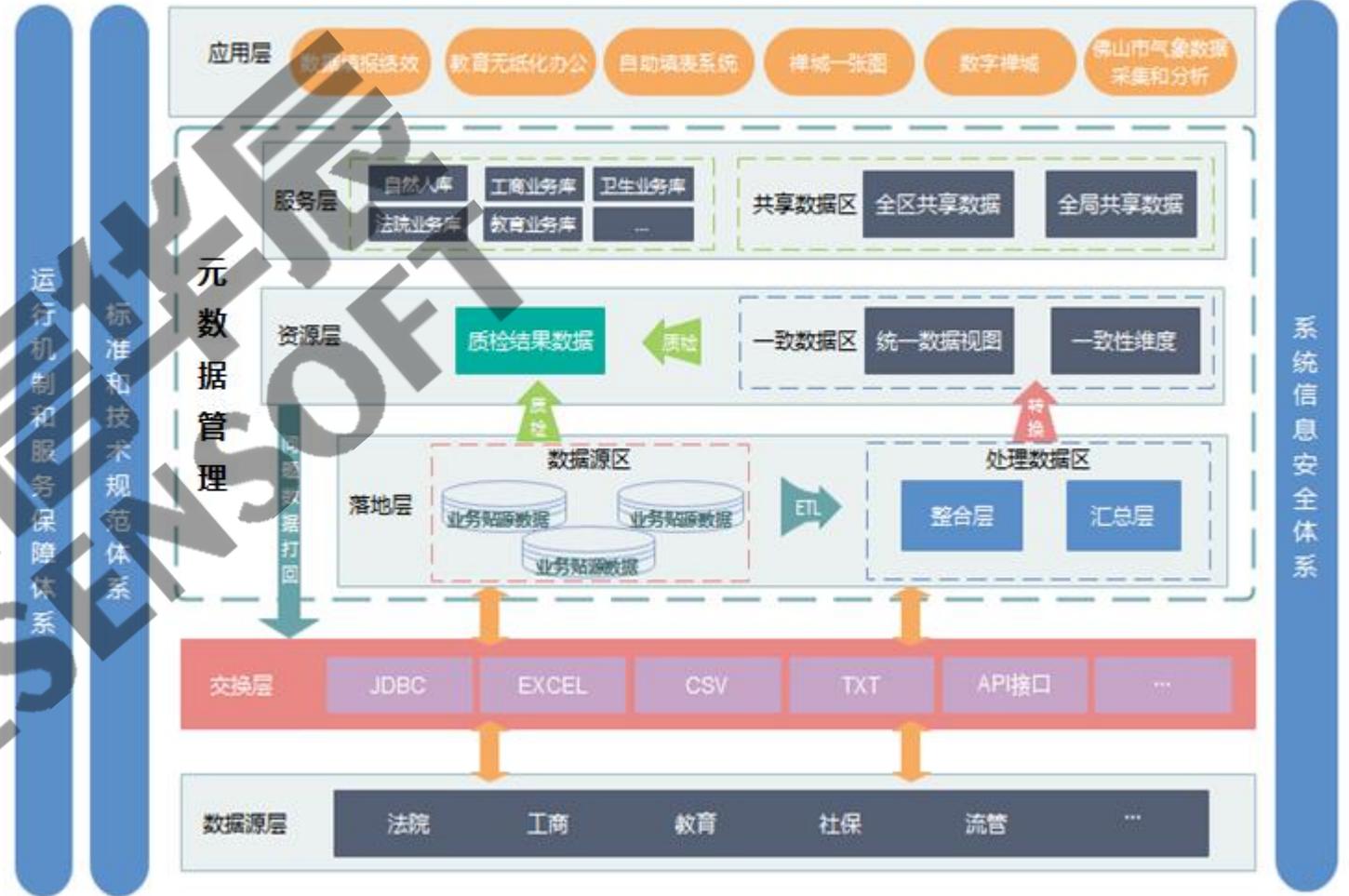
项目概述

2014年3月，启动“一门式”改革，吹响了政务大数据改革的号角；2016年7月，区数据统筹局正式挂牌成立，在区数据统筹局发展决策委员会的指导下，构建“一局、一中心、一公司”的组织体系，推进数统统筹发展的共享、共建、共用，实现“用数据决策、用数据监管、用数据创业”的数据统筹发展运行机制。

区数据统筹局于2018年开展数据治理系统建设工作，通过构建模型、配置规则、人工智能识别等方法对禅城区数据进行数据核查、数据清洗、数据更新及数据修复，并输出数据分析报告，提出改进建议，为该区各个部门数据源提供数据标准规范及改进建议，为政府部门、事业单位、企业、市民等数据使用者提供更优质、更高效的数据服务支持。

元数据在项目中的价值

以治理自然人库数据为主，重点治理卫计、工商、流管三个部门的数据，为部门及数据应用提供高质量数据支撑。采集统筹局落地层、资源层、服务层的元数据信息，提供数据来源信息、数据影响范围等服务，给出错误修改建议，提高填报成功率和后续分析应用的可靠性。

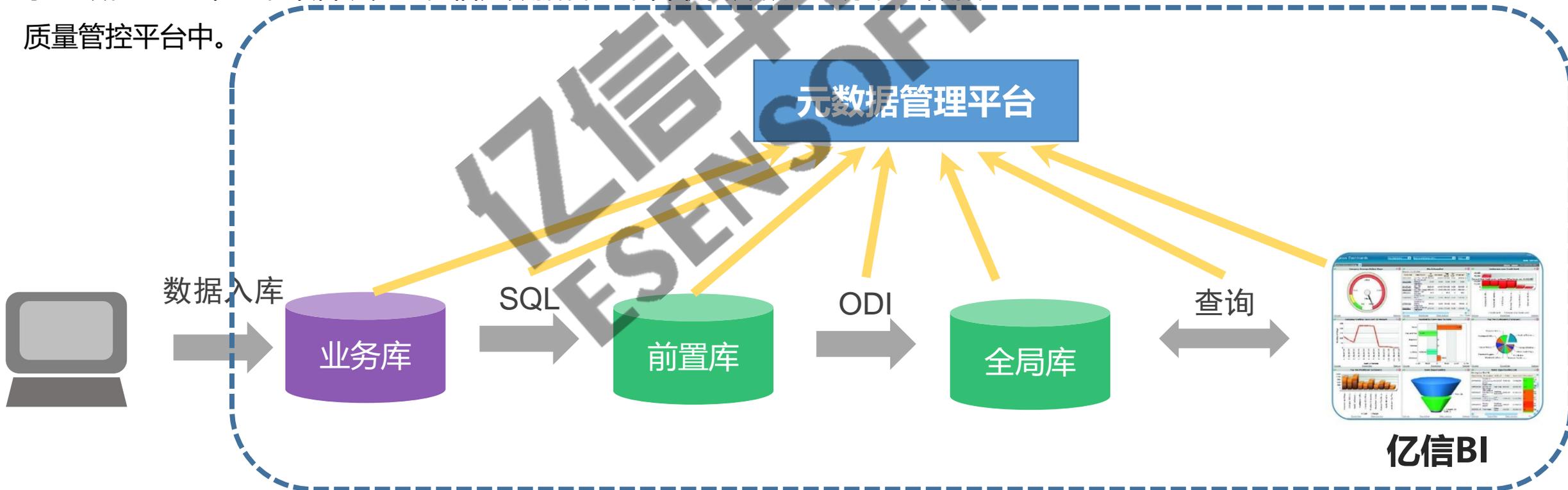


项目概述

某高校在2013年上线的数据质量管控平台，用于管理和监控校内数据质量情况，并出具质量报告。其中包含一个简易的元数据信息采集平台，随着数据不断接入，原系统中采集的元数据信息与元数据相关应用无法满足日益增长的查询和分析需求。故在2018年上半年启动，由亿信元数据管理平台取代并嵌入到原有的数据质量管控平台中。

元数据在项目中的价值

采集和管理全校所有系统相关元数据信息，实现数据从产生，抽取、整合到最后的分析应用的所有过程的可视化集中展现。完善了现有的数据质量管控平台在元数据管理中的薄弱环节，为后续的数据分析、挖掘等深度应用提供可靠的支撑。且形成了教育行业的通用解决方案，在全国多家高校进行了推广。

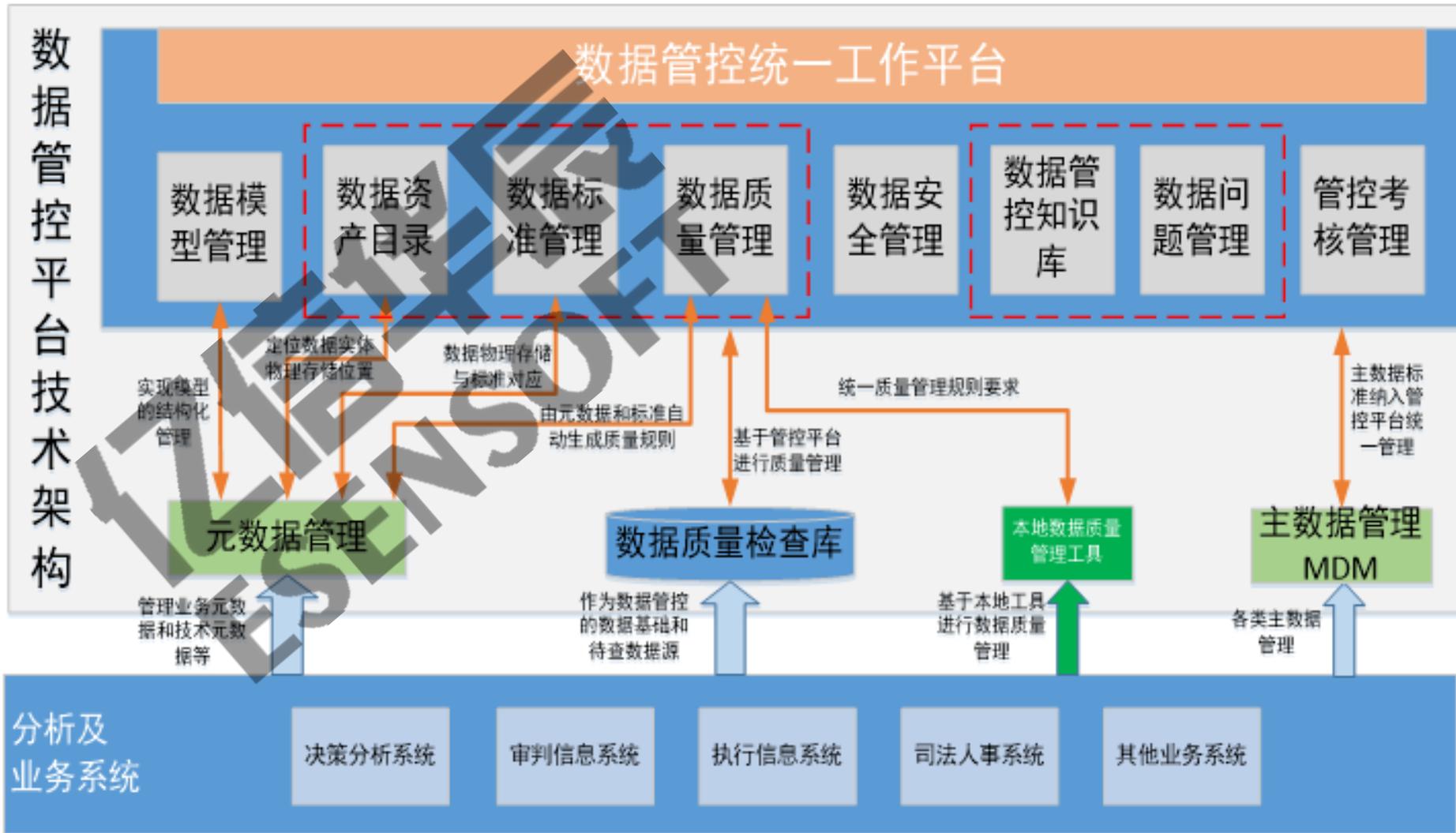


项目概述

为落实国家智慧法院，推进法院信息化相关政策，某法院为充分挖掘信息价值，统一管理数据资产，建设了数据管控平台，用于统一管理与监控现有数据资产，并进行了一些主数据的建设。

元数据在项目中的价值

元数据管理平台实现了业务元数据和技术元数据的集中管理。并为数据模型管理、数据资产目录、数据标准管理、数据质量管理实现无缝对接，提供相关数据、和应用方面的支撑服务。目前该

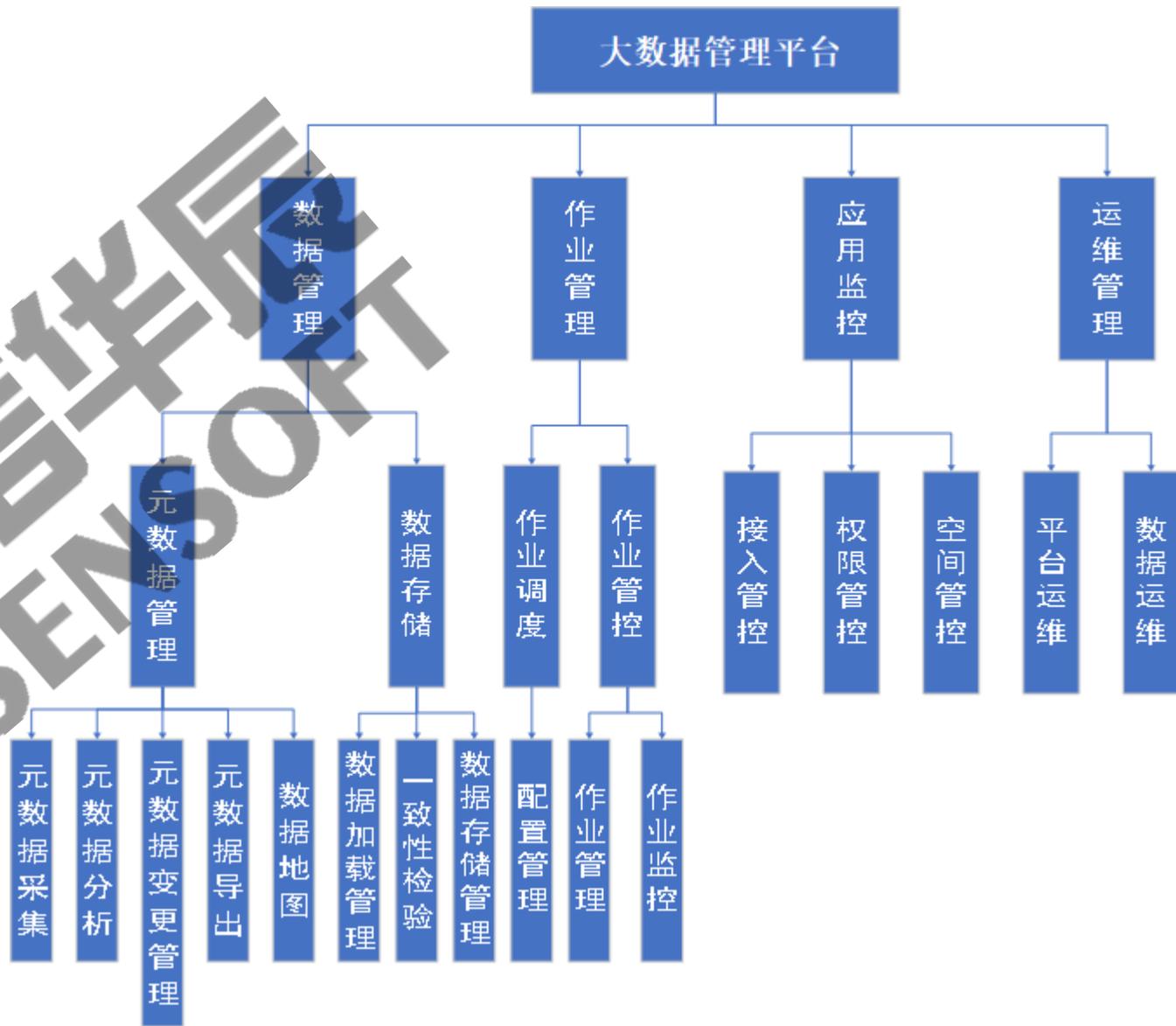


项目概述

大数据平台的总体目标通过整合分行从数据下载系统与数据拆分下发环境获取的数据以及其他行内外数据，打造分行级的统一的、完整的、可信的数据源平台，通过本次扩容引入大数据平台整体的架构规划及系统ETL策略的全局规划，实现数据自动加载入库、ETL脚本自动调度、大数据平台运维整体有效监控等。

元数据在项目中的价值

元数据管理平台主要管理大数据平台中数据与作业的元数据信息，实现元数据信息自动化采集、全局查看数据资产与流向信息。同时提供血缘、影响、关联度等多种分析与元数据的变更管理，协助技术与业务用户进行平台数据的管控工作。





亿信华辰合作伙伴招募计划

携手共赢，开启智能数据新起点



项目合作



产品/服务合作



OEM合作

数据治理专题培训第二期打卡学习心得

1、社区打卡跟帖谈谈你学完第二期课程后，对数据治理和亿信元数据管理产品的理解

帖子入口：<http://bbs.esensoft.com/thread-136240-1-1.html>

打卡模板：#第二期元数据管理——数据治理最基础的一环作业打卡#+本期直播课具体的感受+
你的疑问+其他的想法。



我们的培训绝不止于此

ESENSOFT



ESENSOFT 首页 产品 解决方案 案例 学习 服务 伙伴 关于 社区 400-0011-866

亿信华辰
智能数据全生命周期产品与服务提供商

提供数据采集、数据存储、数据治理、数据分析产品与服务

视频介绍 产品推荐 资料下载

在线客服 联系电话 社区交流

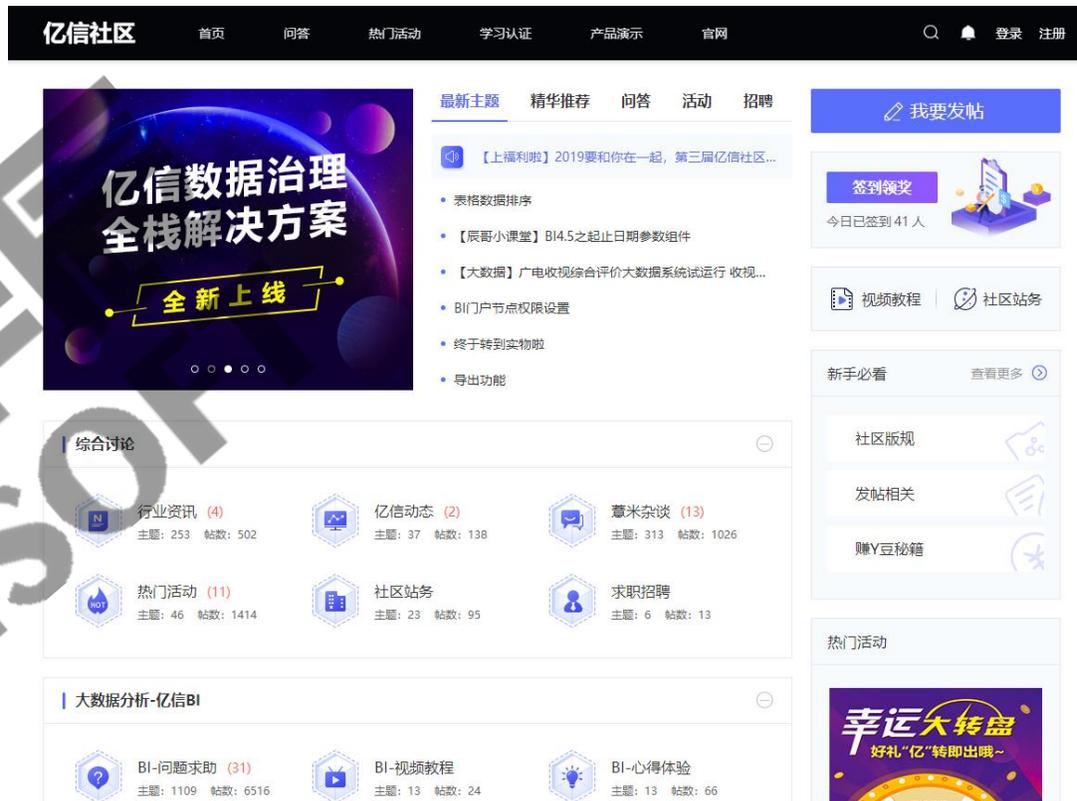
产品更新 PetaBase-s实时大数据平台发布 6大特性拥抱大数据生态	华为认证 亿信BI和查询与华为GaussDB 完成产品兼容性互认证	荣誉奖项 亿信华辰荣登CB Insights 中国企业服务竞争力榜单	数据资产管理白皮书 亿信华辰深度参与编写 为企业提升数据资产价值
---	--	---	---

覆盖数据全生命周期的产品线

深耕大数据各环节，让数据驱动企业变革与进步

 数据存储	 数据处理	 数据治理	 数据分析	 运营监控
---	---	---	---	---

亿信官网



亿信社区 首页 问答 热门活动 学习认证 产品演示 官网 登录 注册

最新主题 精华推荐 问答 活动 招聘

【上福利啦】2019要和你在一起，第三届亿信社区...

- 表格数据排序
- 【辰哥小课堂】BI4.5之起止日期参数组件
- 【大数据】广电收视综合评价大数据系统试运行 收视...
- BI门户节点权限设置
- 终于转到实物啦
- 导出功能

我要发帖

签到领奖
今日已签到 41人

视频教程 社区站务

新手必看 查看更多

社区版规 发帖相关 赚Y豆秘籍

热门活动

幸运大转盘
好礼“亿”转即出哦~

综合讨论

 行业资讯 (4) 主题: 253 帖数: 502	 亿信动态 (2) 主题: 37 帖数: 138	 薏米杂谈 (13) 主题: 313 帖数: 1026
 热门活动 (11) 主题: 46 帖数: 1414	 社区站务 主题: 23 帖数: 95	 求职招聘 主题: 6 帖数: 13

大数据分析-亿信BI

 BI-问题求助 (31) 主题: 1109 帖数: 6516	 BI-视频教程 主题: 13 帖数: 24	 BI-心得体验 主题: 13 帖数: 66
--	---	---

亿信社区

我们的培训绝不止于此

ESENSOFT



亿信官微



ESENSOFT

感谢您的参与

Thank you for your participation

智能数据全生命周期产品与服务提供商

