

Real-time lakehouse integrated data development platform

实时湖仓一体化 数据开发平台

助力企业数字化转型

实现数据价值变现且持续增值



目录

CONTENTS

01

数字化转型需求

Digital transformation needs

02

产品方案介绍

Product solution introduction

03

数据湖架构介绍

Platform function display

04

平台功能展示

Platform function display



PART 01

数字化转型需求

Digital transformation needs





数字化转型面临的数据挑战

数据“用不好”挑战

数据量越来越大，种类及业务场景多样化、复杂化。对时效性要求越来越接近在线。对数据资产价值认识不足缺乏大数据应用经验，应用的规划局限打，应用建设成效不理想；

数据“拿不到”挑战

跨部门、跨组织、异构数据难以直接共享开放，缺乏完善的交换体系和机制。

数据“不知道”挑战

1) 数据在不同部门相互独立存储，独立维护，彼此间相互孤立，形成“孤岛”，难以在数据源端实现数据的互联互通。2) 不同部门对数据进行理解和定义不同的含义，无形中加大了跨部门数据合作的沟通成本，业务系统改造升级难度大

数据“管不住”挑战

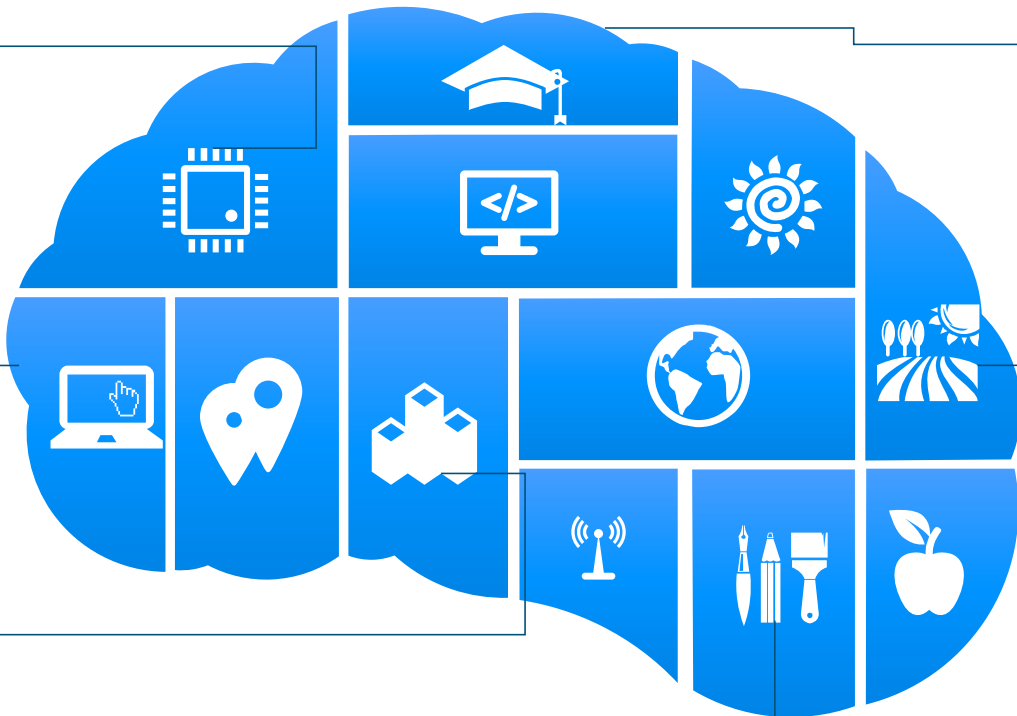
随业务覆盖更完整，总量、类型、数据源、历史深度急剧增长。数据怎么管，数据质量无法管控，数据安全性无法保证。信息不全面、不及时、不准确，容易造成管理漏洞

数据“两张皮”挑战

数据做生产资料，其中数据分析成为核心业务功能，不同场景下的业务服务需要多样化支撑、融入生态，许多标准并没有真正落实到管理，体现在实践中。只有在外审时才能暴露出来。

数据“夹生饭”挑战

数据湖、数仓、湖仓一体、数据中台等企业统一数据支撑平台涉及到多种技术栈路线，如关系型数据库、Hadoop、MPP、时序数据库、文档数据库、图数据库、容器、K8S等技术栈，加大了客户侧统一集成的难度。而且，很多客户并未真正的将这些技术用起来



企业在向数字化运营的转型过程中的数据挑战

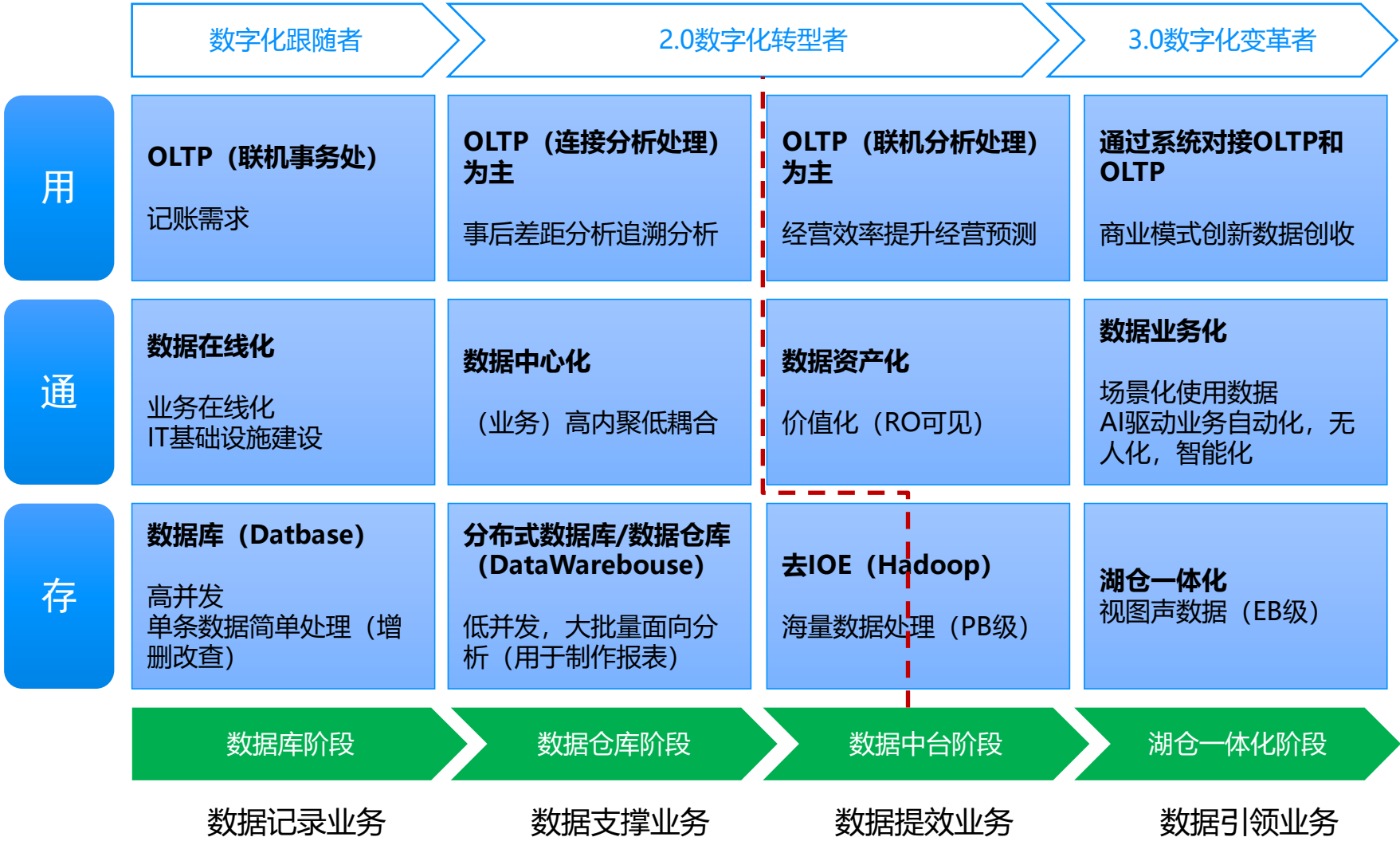
需要打造一个的统一兼容好数据底座，为多样场景下的业务赋能提供有效支撑



数字化建设发展阶段

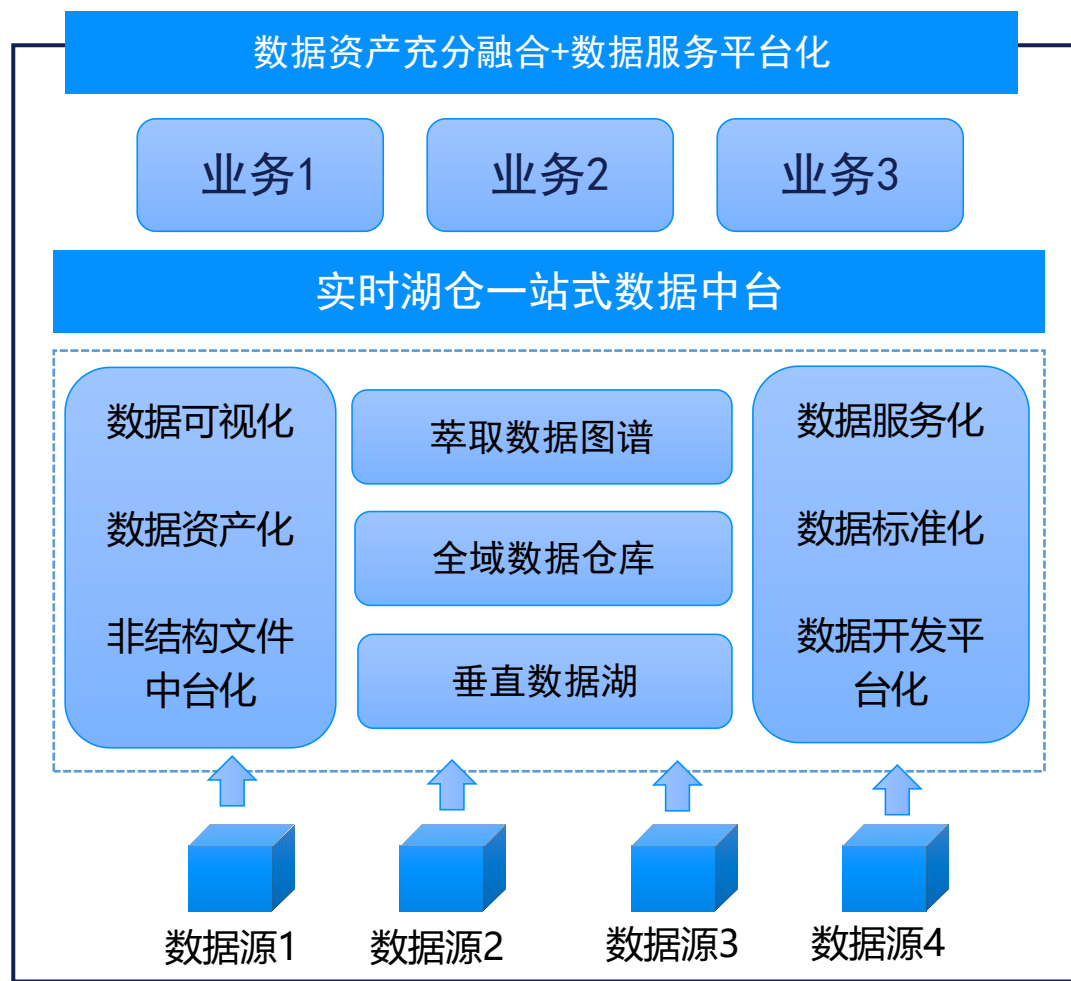
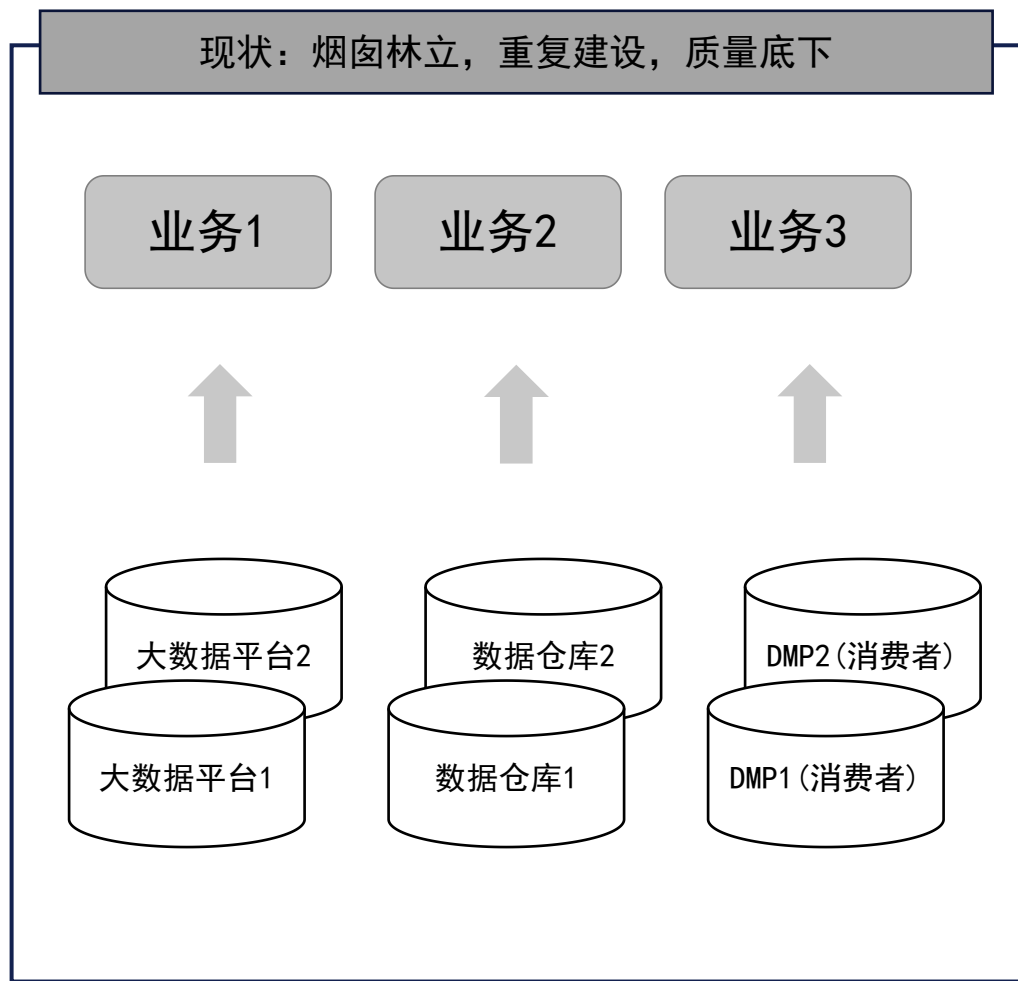
★ 企业所处阶段

☆ 企业未来目标



- **数据库阶段**
主要是OLTP (联机事务处理) 的需求
- **数据仓库阶段**
数据仓库阶段, OLAP (联机分析处理) 成为主要需求
- **数据中台阶段**
主要解决BI和报表需求的技术问题
- **湖仓一体化阶段**
通过系统来对接OLTP (事务处理) 和OLAP (报表分析) 的需求, 强调数据业务化的能力

大数据的技术，理念、方法论都需要升级





PART 02

产品方案介绍

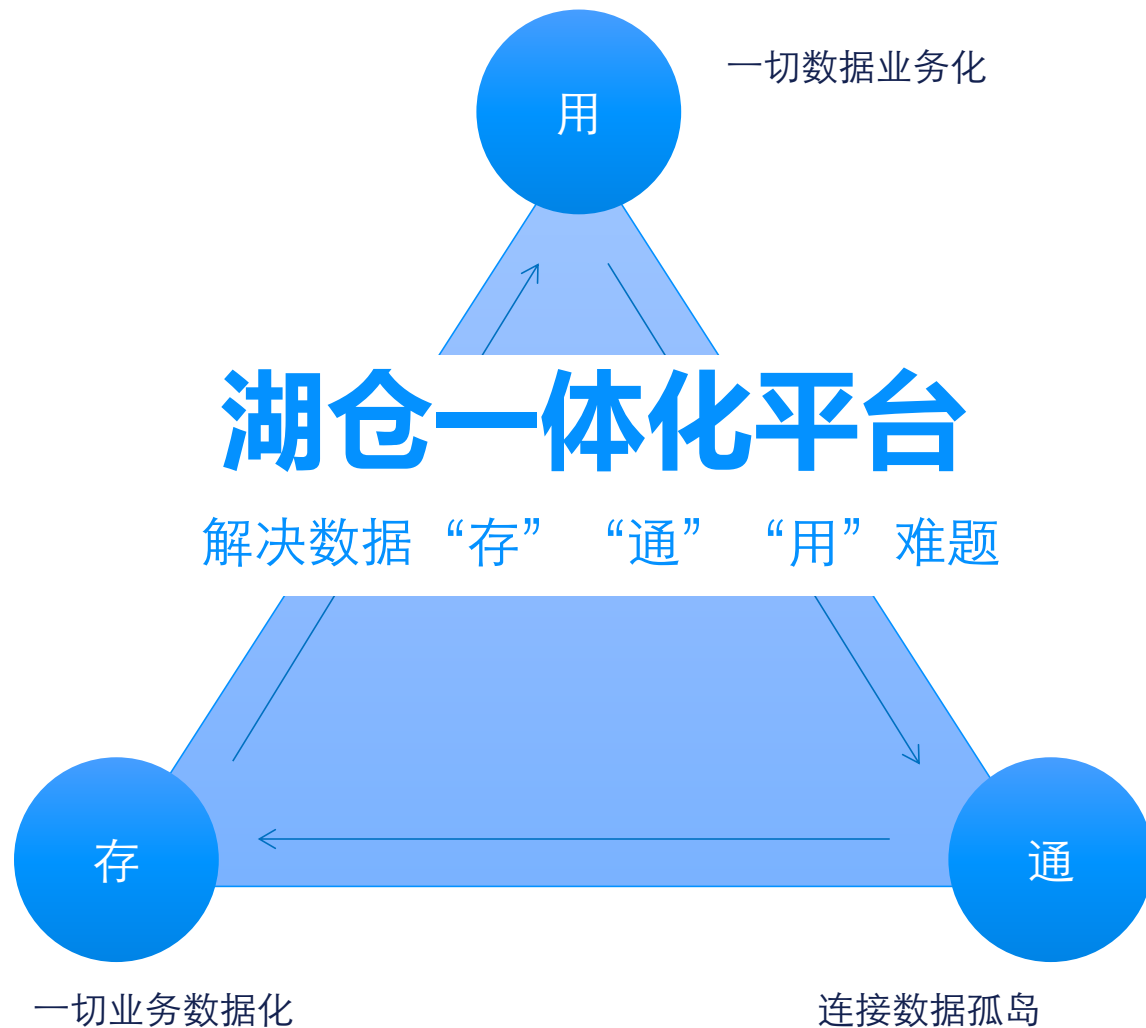
Product solution introduction





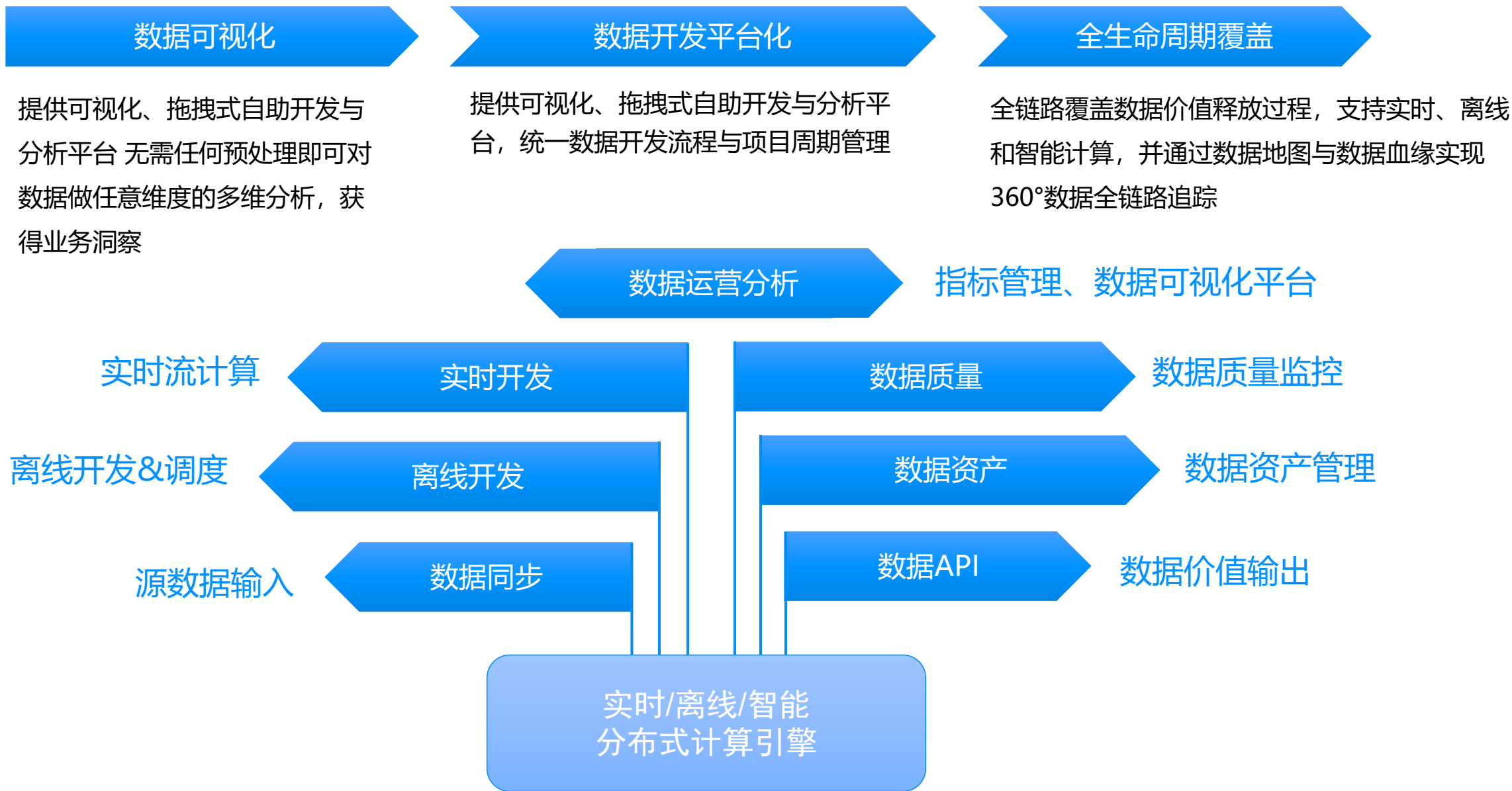
产品定位：让数据产生价值

产品理念是让企业的整个业务都可以共享同一套数据技术与资产，平台中的数据源各自业务系统，为业务体系提供有利的算法、数据和技术支持。构建统一数据治理体系，实现数据资产运营有标准指导执行、有专业平台承载，持续优化数据资产的管理流程及体系，围绕数据采集、整理、分析、挖掘、展现及应用等环节，打造一存储管理、分析挖掘及可视化为核心的动态、可伸缩、可复用的优质数据服务能力。之后通过统一化的数据技术服务反哺业务。





产品价值：让大数据从“DOS时代”进入“Windows时代”



平台矩阵



数据使用部门

统一数据服务、统一数据分析



数据使用部门

统一数据服务、统一数据分析



数据管理部门

统一数据资产管理



数据开发部门

统一数据集成、数据治理



数据架构部门

数据平台存储引擎

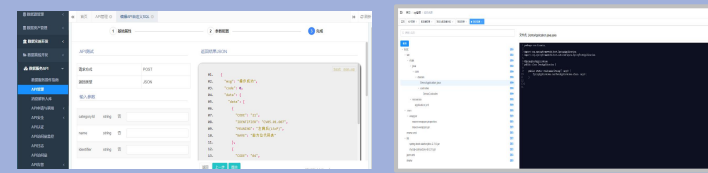
数据服务平台



可视化分析

数据指标

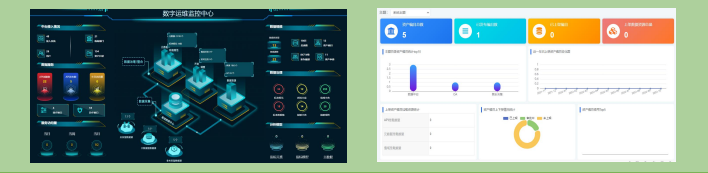
数据共享分析平台



数据服务

文件中台

数据资产平台



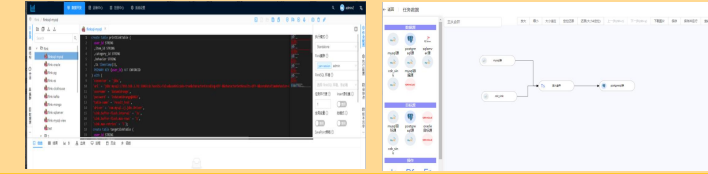
资产盘点/数据
生命周期

数仓建模/数据
脱敏

数据标准/数据
质量

主数据/元数据/
数据地图

数据开发平台



数据实时开发

数据离线开发

数据存储底座

湖仓一体

实时数仓

离线数仓

数据湖

一湖

- 数据湖

两仓

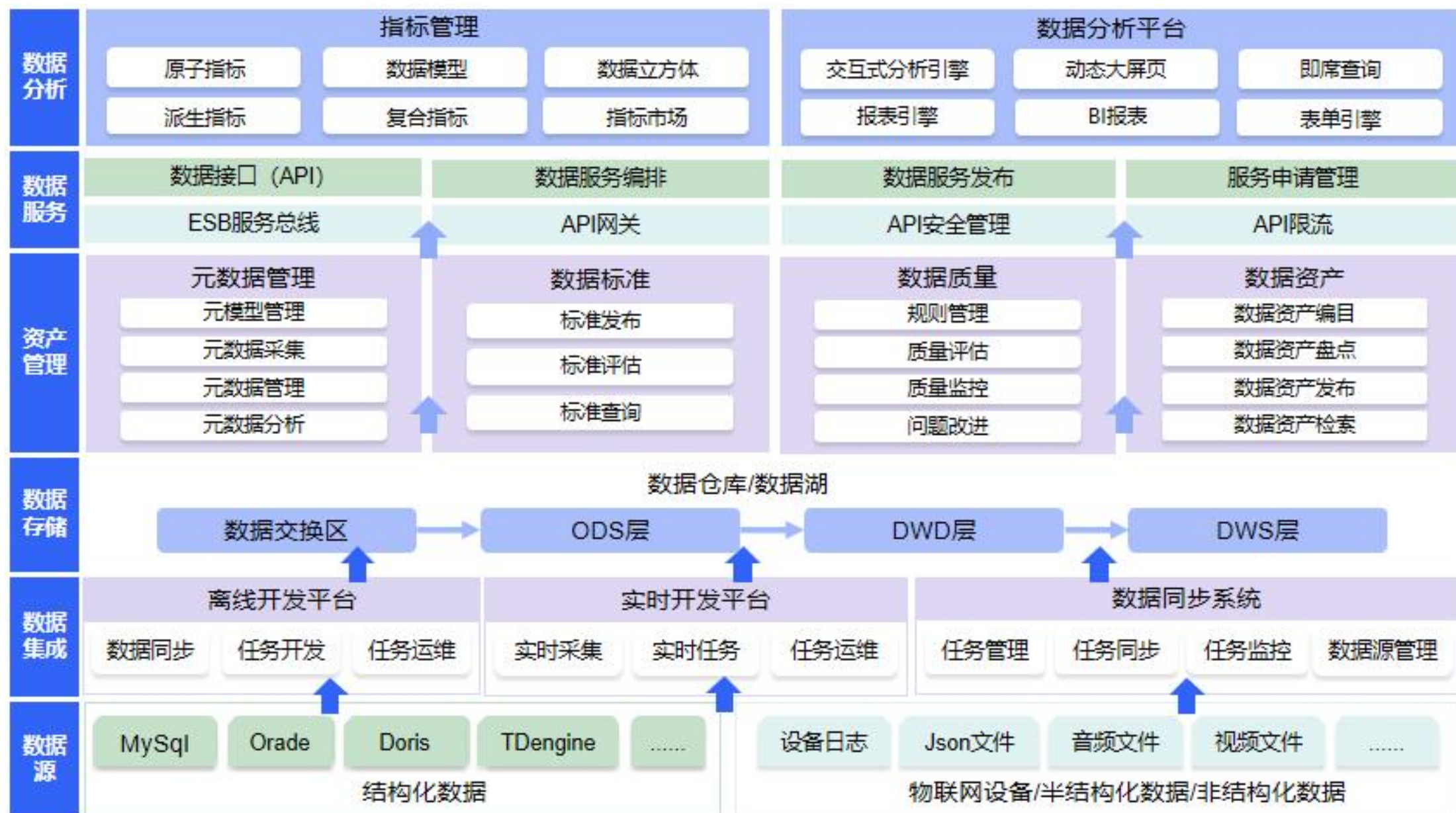
- 实时数仓
- 离线数仓

四平台

- 数据分析平台
- 数据服务平台
- 数据开发平台
- 数据资产平台



平台功能架构





平台核心功能

数据服务平台



数据服务发布中心，可以将数据中台的数据共享给其他系统，无缝对接业务生产系统，主要解决数据对外快速共享的场景

数据同步系统



对异构数据源实时高性能的一站式数据同步系统

数据质量管理



通过向导、可视化的简易操作手段，将质量评估、质量检核与质量报告等工作环节进行流程整合，形成完整的数据质量管理闭环

数据资产管理



企业数据资产门户，全域数据资产地图，在识别出自身数据资产的基础上，进一步构建数据资产目录，帮助用户更好的理解、使用以及分析数据

数据指标管理



一站式企业数据指标管理平台，构建企业数据指标体系，落地指标数据结果，消除数据二义性，实现指标数据的可视、可用、可管

数据可视化



提供统一的自助式数据分析和决策辅助功能，将之前单一的数据进行多维度的关联分析

数据开发管理平台



一站式数据开发管理平台，可实现数据离线和实时开发。降低企业大数据开发的技术门槛，提升数据开发效率

数据安全



基于数据中台的数据脱敏管理系统，能够有效的将客户的敏感信息进行保护，避免了敏感信息在进行使用的时候发生泄露，保障数据的使用隐私

文件中台



实现企业非结构化数据的统一存储、统一管理和统一服务，建立统一的非结构化数据管理平台，实现数据互通、业务融合



产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

数据建模

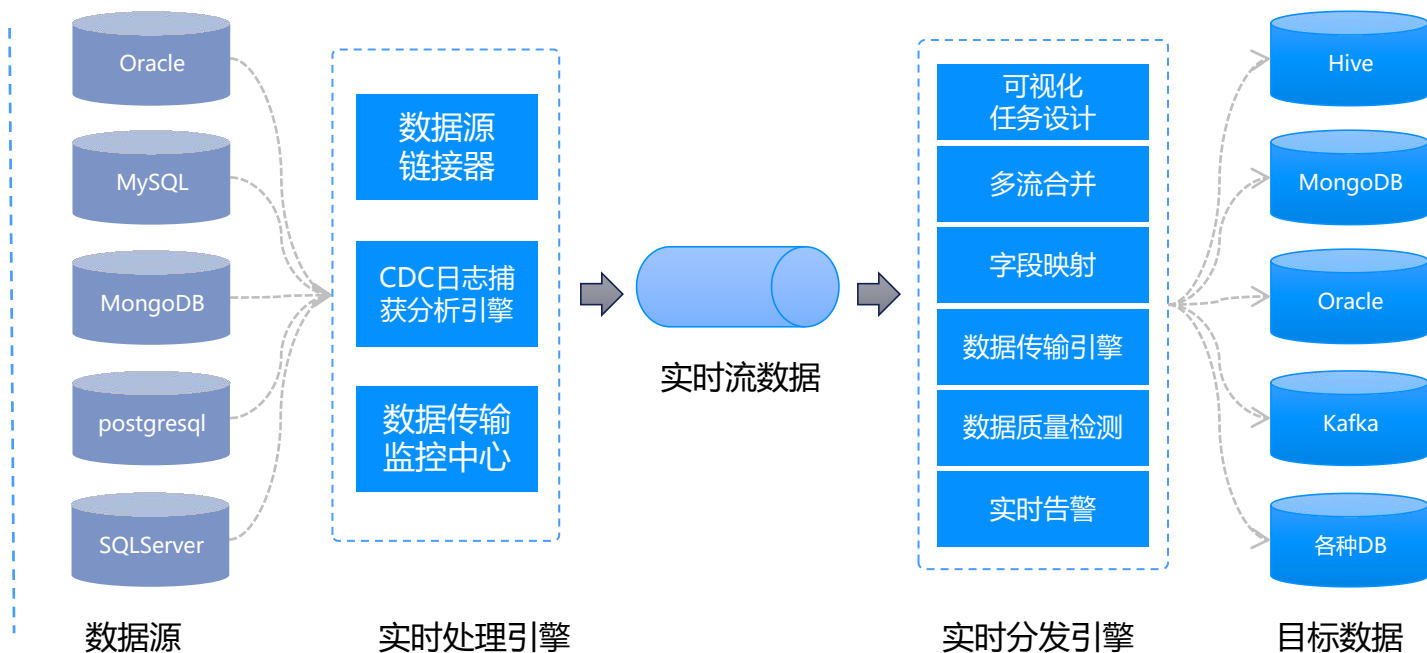
数据资产

数据质量

数据服务

数据运维监控

- 实时数据处理、集成、分发
- CDC数据变更捕获技术
- 增量数据实时同步、处理
- 变更数据实时发送



- 数据同步底层是基于Flink自研的一套分布式架构，支持大容量高并发同步，支持同步速率和脏数据管理
 - 支持多种不同的异构存储系统数据，插件化架构可随时支持更多的新数据需求，支持全量、增量方式
- ① **数据库类:** MySQL/Oracle/SQLServer/PostgreSQL/瀚高/达梦/金仓/FTP...
 - ② **NOSQL类:** Redis/MongoDB/Hbase/ElasticSearch...
 - ③ **大数据类:** clickhouse/HDFS/Hive/Doris(MPP)...



产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

数据建模

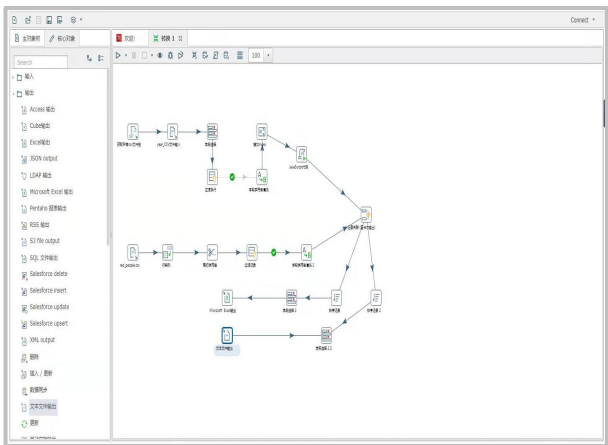
数据资产

数据质量

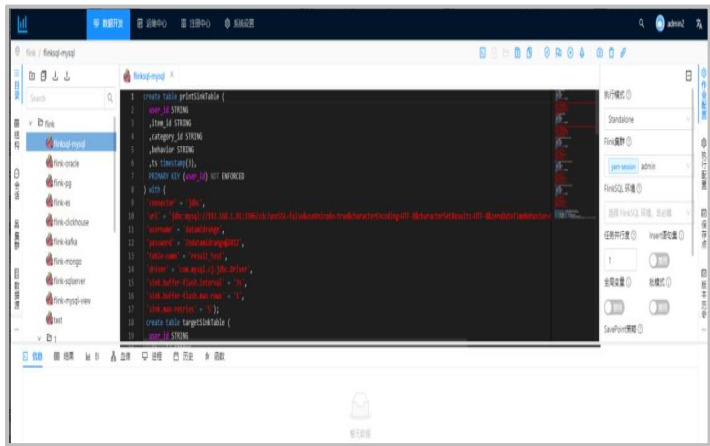
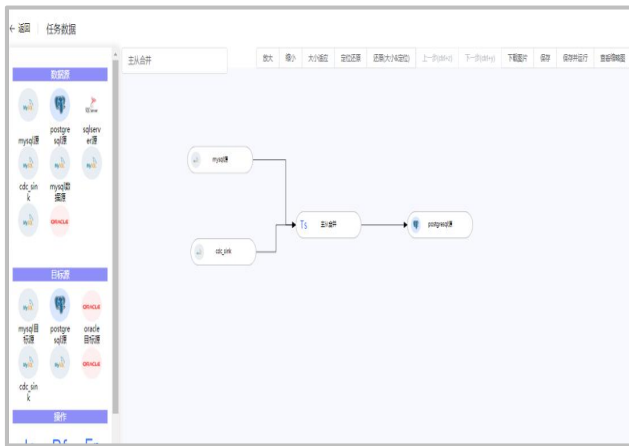
数据服务

数据运维监控

离线开发



实时开发



• 离线任务研发

便捷的ETL流程设计器：全程拖拽式、易操作、易阅读、易维护

丰富的数据处理组件：50余种数据处理组件、30余种清洗规则

• 实时任务研发

任务类型：Flink、FlinSQL、实时采集

任务配置：源表支持Kafka、，结果表支持MySQL、Oracle、Hbase、ES、Redis、MongDB、Kafka、维表支持MySQL、Oracle、Hbase、Redis、MongDB

功能特点：非侵入式数据同步、多任务并行和高并发、纯国产自研、版本管理和任务锁机制



产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

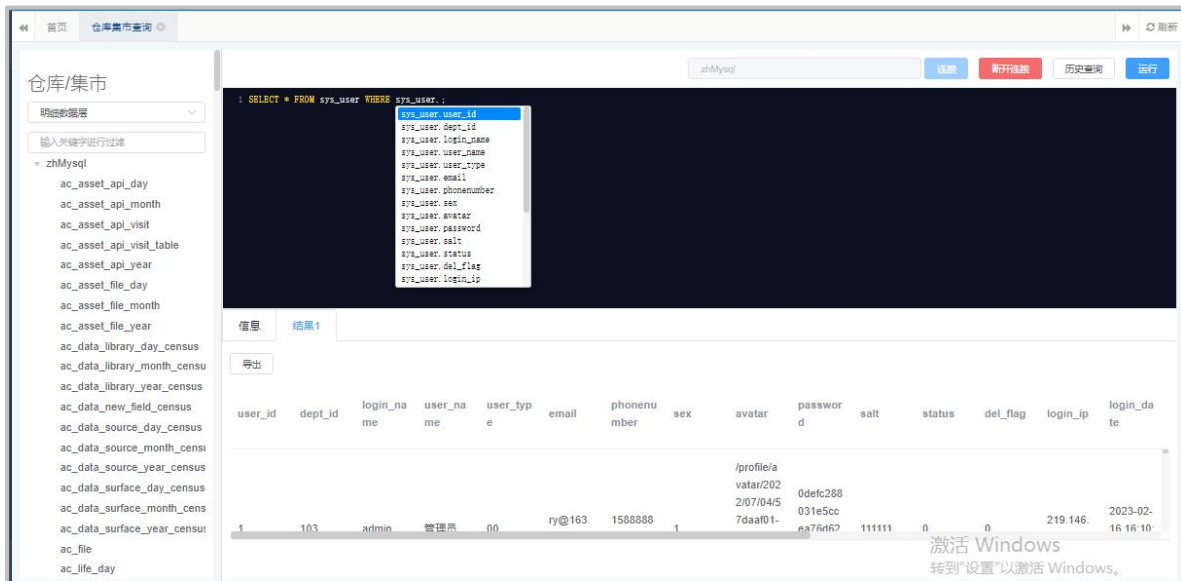
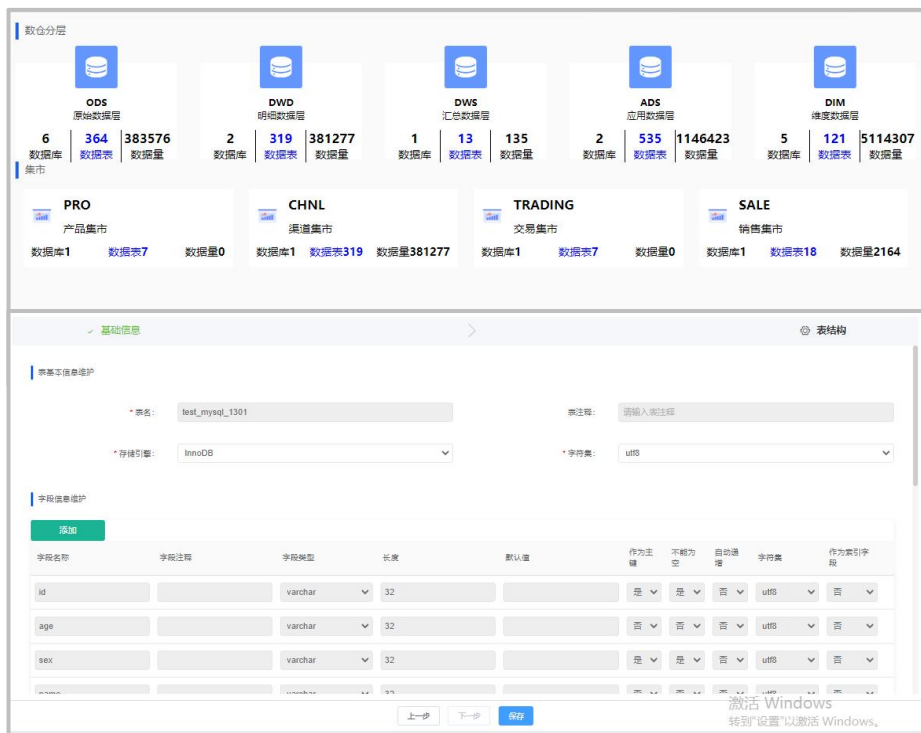
数据建模

数据资产

数据质量

数据服务

数据运维监控



- **数仓分层设计：**提供可视化的界面操作功能，实现数据仓库规划设计，实现面向业务、面向分析的各类主题域划分，使数仓的设计变的简洁方便清晰。
- **数据建模：**支持逻辑模型、物理模型设计，保障建模流程规范可控，支持MySQL、Oracle、SQLServer、Drois、Hive、达梦、金仓、Tidb
- **数据查询：**提供界面化查询入口供数据分析人员对数据中心中的数据进行查询分析与探索。



产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

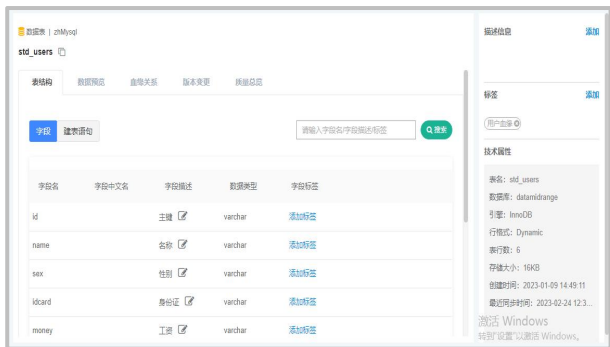
数据建模

数据资产

数据质量

数据服务

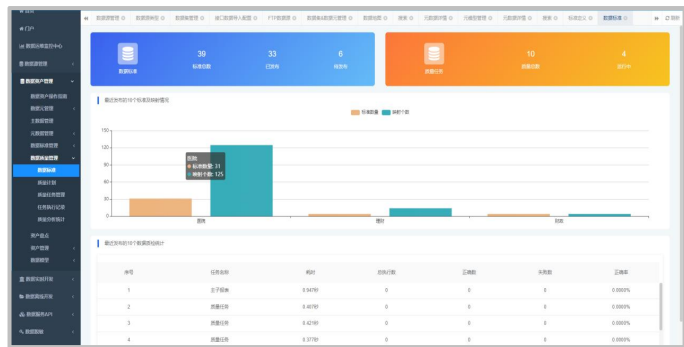
数据运维监控



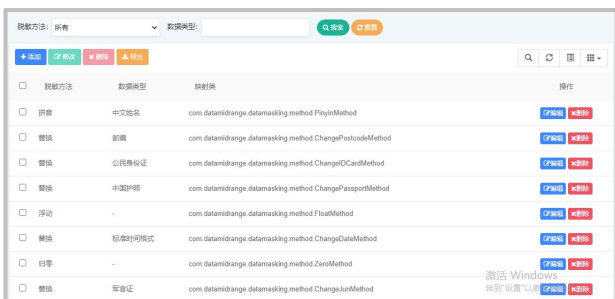
元数据管理



血缘管理



数据质量



数据脱敏



数据标准



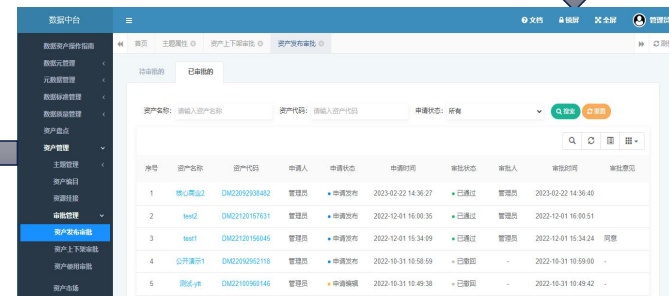
数据资产

对所有资产数据进行统一管理，了解平台的核心数据资产，提供对应数据资产管理规范。资产使用者可快速对资产进行查找、申请、使用等操作。

- **数据血缘：**自动解析同步任务和sql代码，自动建立各个数据表的表级、字段级血缘关系，实现数据的统一管理和血缘的跟踪
- **数据资产：**构建数据资产地图，在识别出自身数据资产的基础上，进一步构建数据资产目录，帮助用户更好的理解、使用以及分析数据。
- **数据脱敏：**基于数据中台的数据脱敏管理系统，能有效将客户敏感信息进行保护，避免敏感信息在使用时发生泄露，系统采用高性能、高扩展性的敏感数据自动发现动态数据脱敏和静态数据脱敏功能。



数据运维监控



资产发布/使用审核

资产市场：已发布并审核通过的各类数据资产，会在资产市场中进行展示，在资产市场中当前登录人可以查看具有权限访问主题和类目下已上架的各类资产编目，并对相应的资产编目进行使用申请



产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

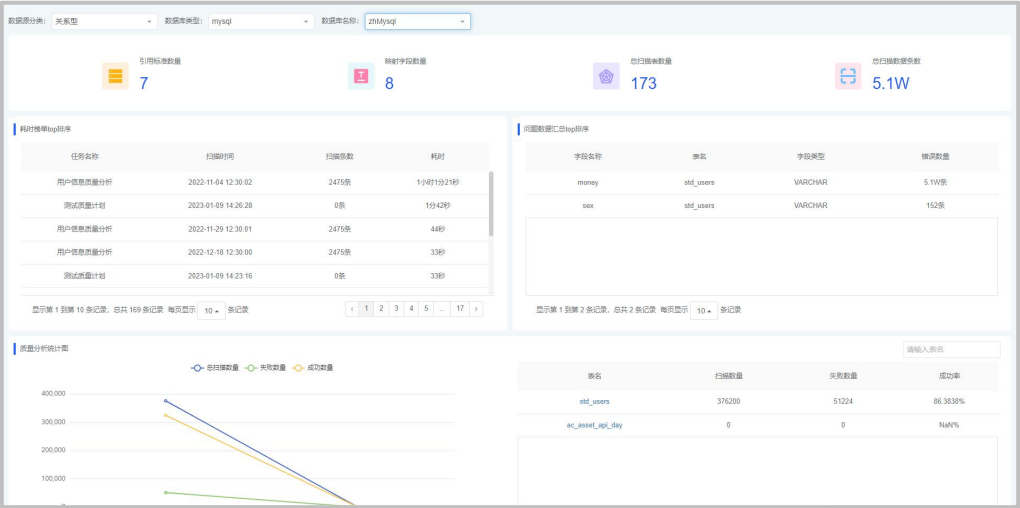
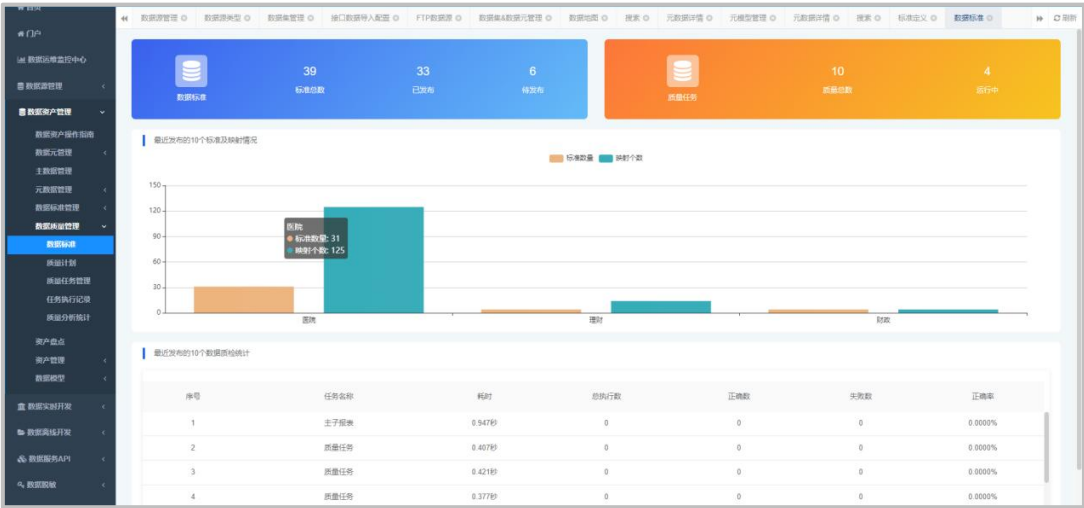
数据建模

数据资产

数据质量

数据服务

数据运维监控



实现数据质量问题的全面筛选和精准定位，以可视化方式实现对数据质量的全局掌控和管理。以数据标准为数据检核依据，以元数据为数据检核对象，形成完整的数据质量管理闭环。

作为数据治理的一部分，数据质量治理包括及时性、完整性、一致性、准确性，落地到具体的平台功能点。

- **支持的数据源：**MySQL、Oracle、PgSQL、Drois、Tidb等
- **支持的校验规则：**表级、字段级、字段格式和自定义sql4类，同一数据源的双表逐行校验规则的配置
- **支持的调度机制：**小时、日、周、月规则调度
- **质量监控：**字段级监控报告和表级监控报告



产品特点：一站式数据开发（7步完成数据中心建设）

可视化“零代码”生成API，3分钟让数据快速服务化，提供开箱即用的API服务能力和轻量级ESB服务总线，精细化的数据安全管控，提高数据开放与共享效率。

数据引入

数据研发

数据建模

数据资产

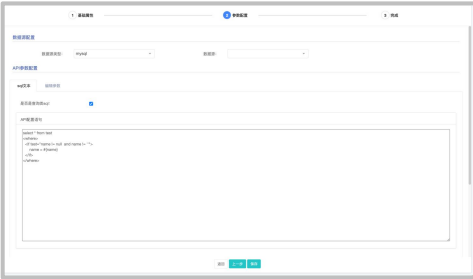
数据质量

数据服务

数据运维监控

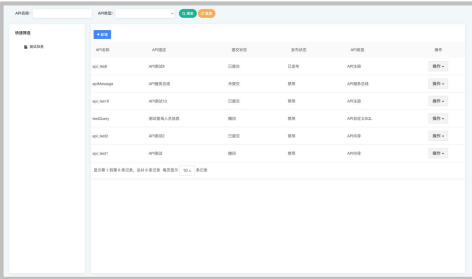
API创建

- 支持20多种数据库创建API
- 支持单表和夺标创建API
- 支持其他平台API注册
- 支持API服务编排式创建



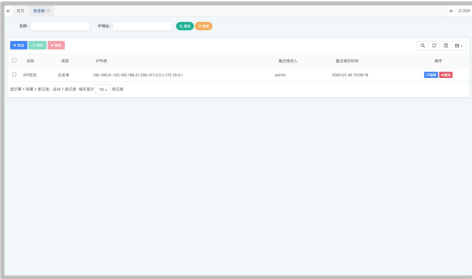
API管理

- API告警配置管理
- API标签目录分类管理
- 支持API调用限流管理
- 支持API停用和启用



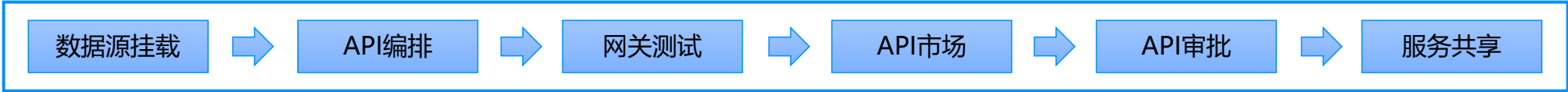
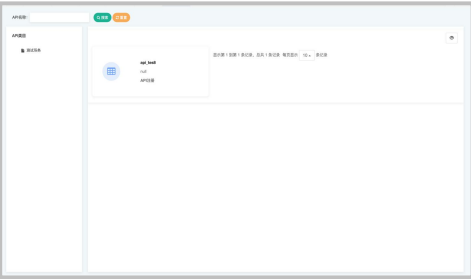
API安全

- API日志审计管理
- 支持API安全组访问
- 支持API行级权限配置
- 支持2种API加密方式



API市场

- 支持多重角色视角查看
- 支持API市场搜索查询
- 可形成企业/部门级API市场
- 支持API导入导出





产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

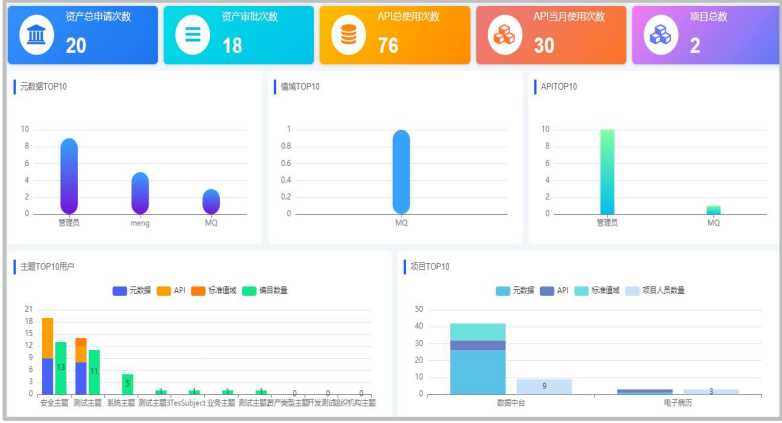
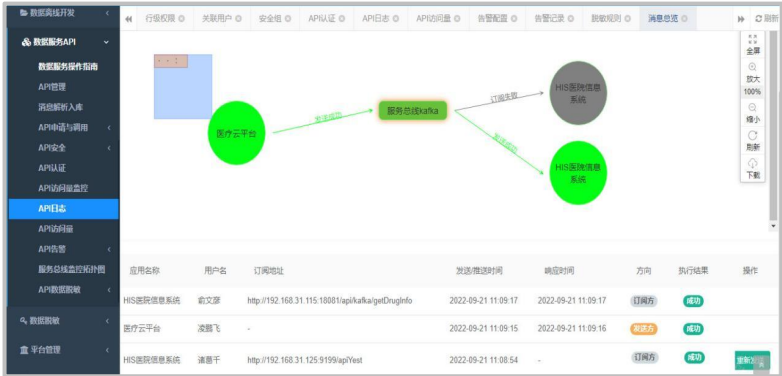
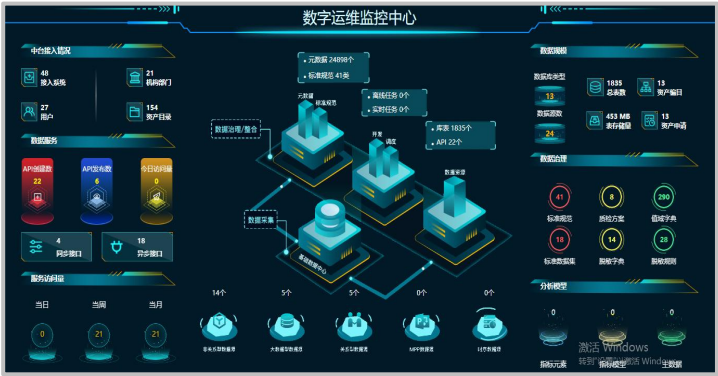
数据建模

数据资产

数据质量

数据服务

数据运维监控



- ❑ **数据运维中心**：在整体的数字运维监控中心，可以查看（元数据、主数据、数据标准、接入系统数量、数据源数量、表存储、数据服务数量、离线任务、实时任务）等平台全局的数据运营情况，
- ❑ **数据资产数据地图**：在元数据基础上提供的企业数据目录管理模块，涵盖全局数据检索、元数据详情查看、数据预览、数据血缘和数据类目管理等功能。可以更好地查找、理解和使用数据。
- ❑ **API访问量监控**：可以实时监控数据服务每日、每周、每月的访问量，查看日志的异常情况，平时响应时间



产品特点：一站式数据开发（7步完成数据中心建设）

数据引入

数据研发

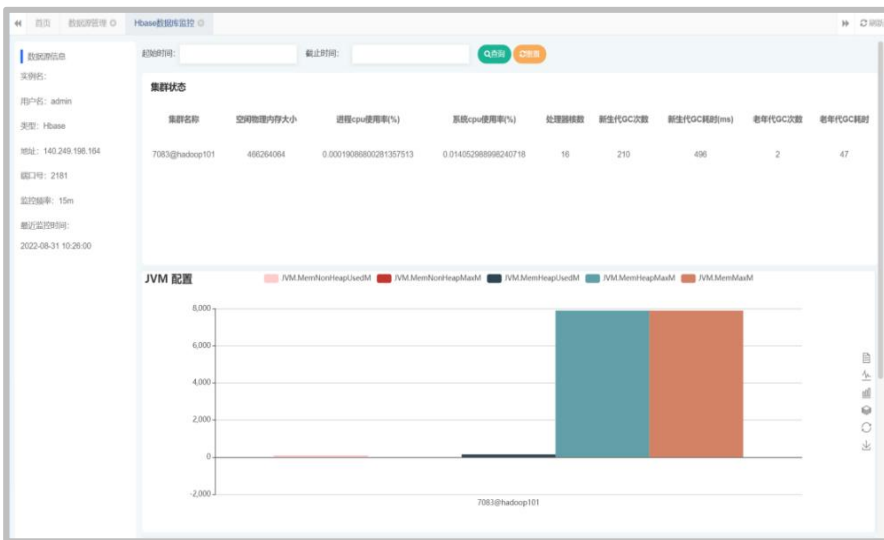
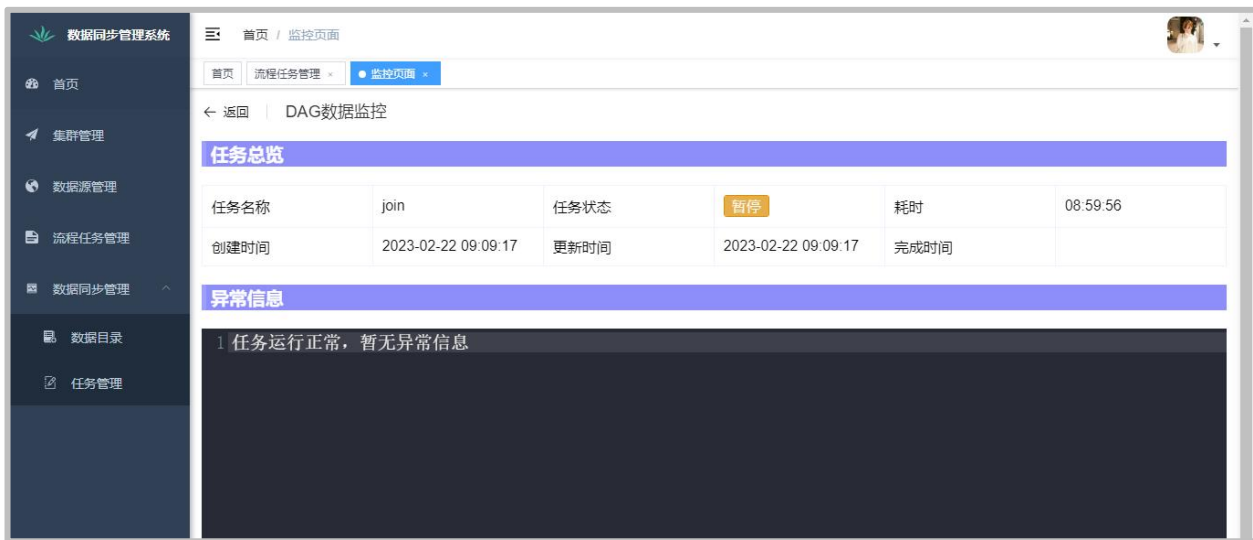
数据建模

数据资产

数据质量

数据服务

数据运维监控



- **任务监控告警：**根据监控规则和任务运行情况，智能监控决策是否报警、何时报警、如何报警以及给谁报警。智能监控可以筛选关键依赖，最终自动选择最合理的报警时间/报警方式以及报警对象，辅助您的决策优化
- **数据源监控告警：**支持对各数据源进行性能指标的监控，查看数据表存储、索引大小、连接检测、流量检测、锁检测、缓存检测、慢查询数量等，实时掌握各数据源的运行状态。



PART 03

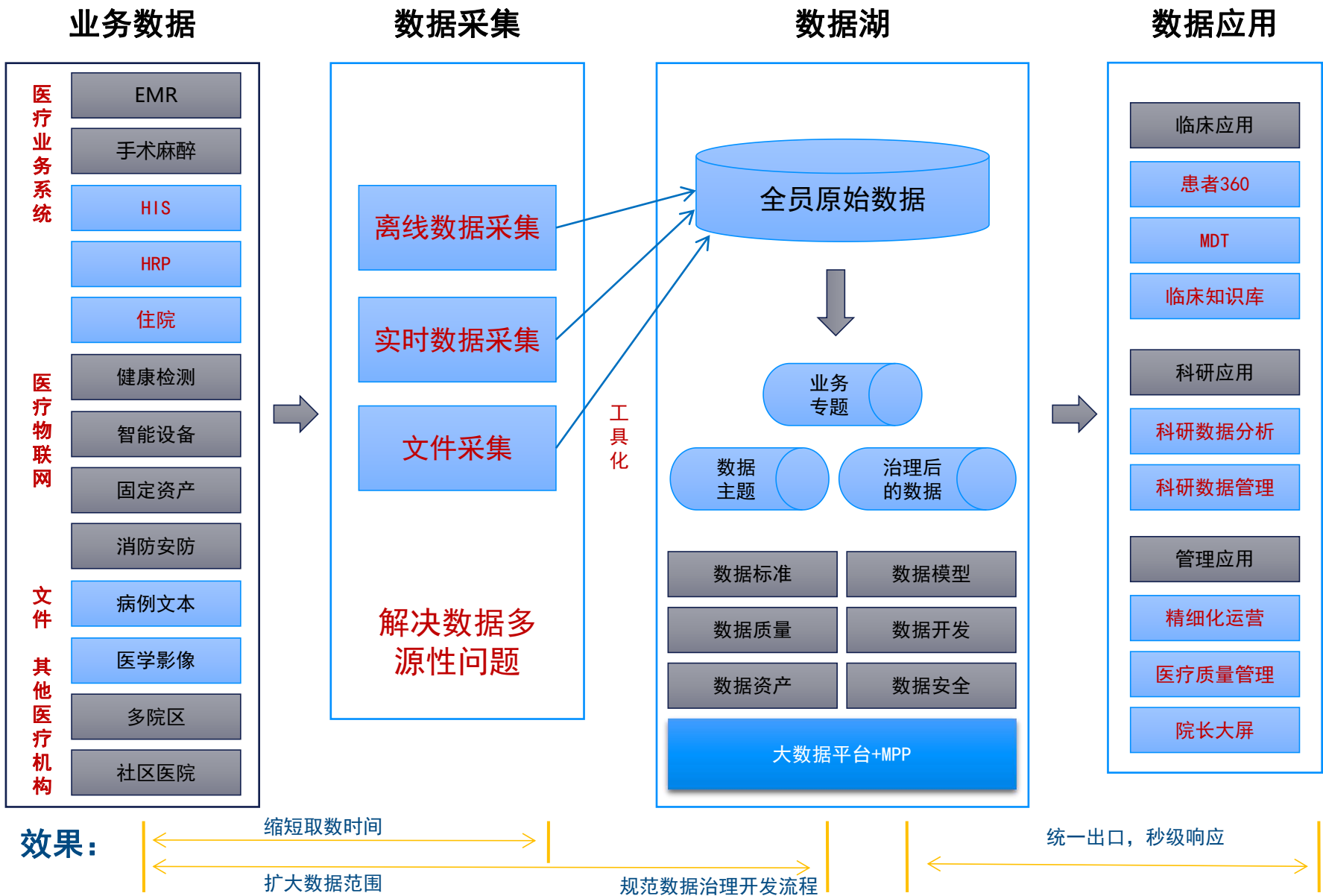
数据湖架构介绍

Platform function display





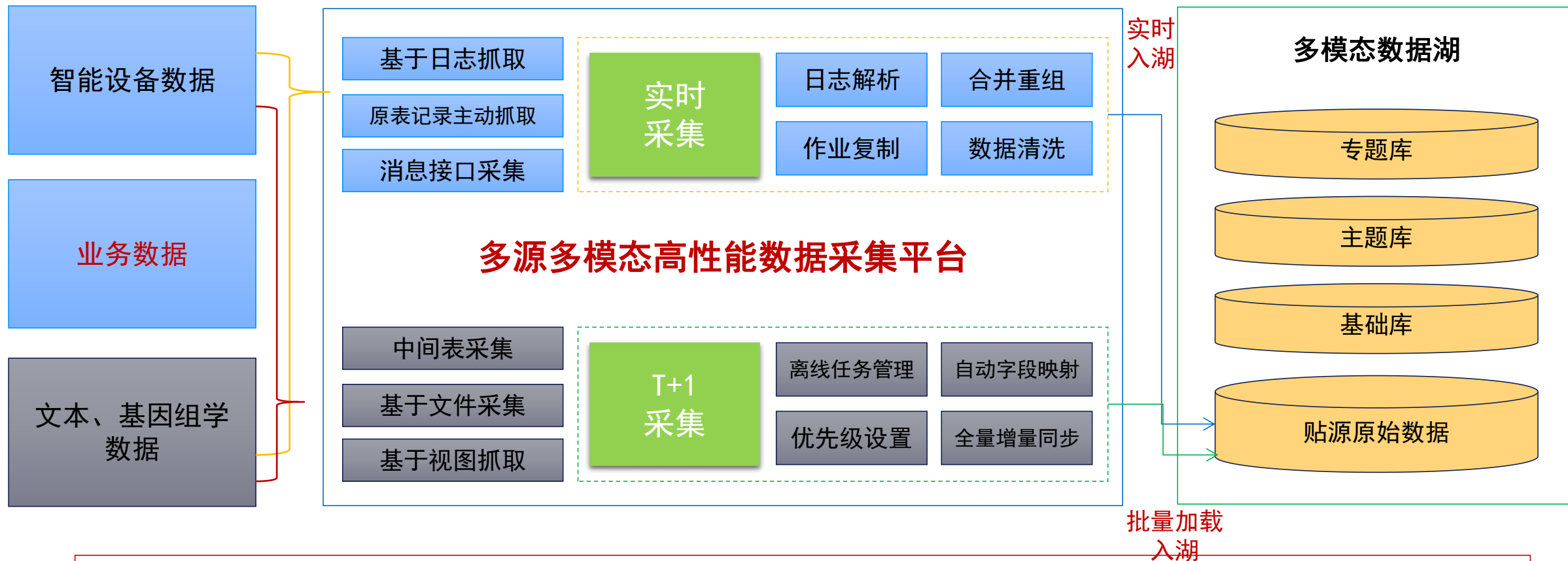
全域数据湖数据架构



架构特点：工具化、自动化、 标准化、高性能

- 不影响业务系统和平台交互：对业务系统侵入性小
- 工具集成和自动化：基于数据湖的ETL、CDC采集可通过湖管理工具完成统一标化处理
- 治理可视化、溯源链路清晰：作为全量数据的对接方，可协助数据治理方找到原始数据的内容及出处
- 沉淀数据资产提升复用能力：数据治理成果沉淀为可复用的数据资产
- 接口易开发、治理结果易复用：避免频繁的从业务系统厂商向下游应用厂商的频繁接口提供

数据入湖：兼容多源多态、保障性能高效



- 量多两优：数据库支持类型多、医疗数据支持种类多
- 性能优化：针对患者主诉大段文本以及心电图、PDF等数据，大幅优化提升数据入湖性能
- 实时高效：支撑数十亿数据采集规模，满足大批量即席数据分析需求基础上更能满足实时业务需求

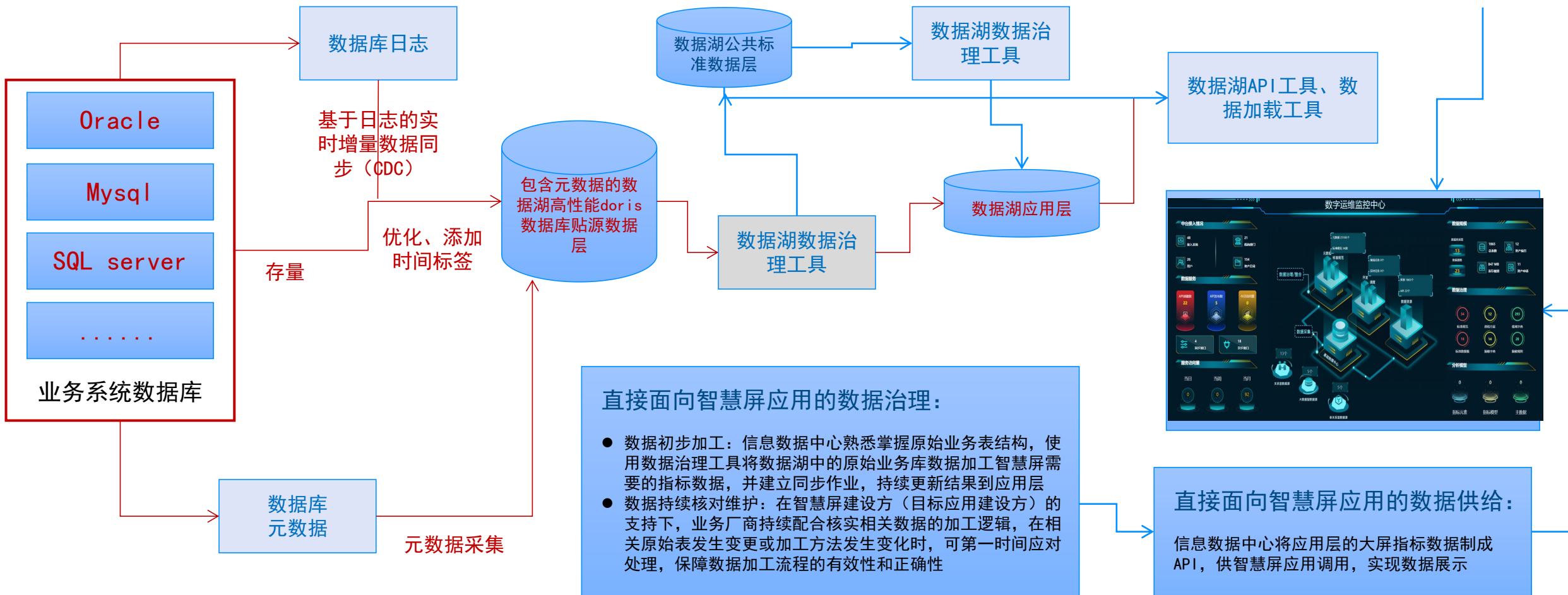


基于数据湖的数据开发流程

基于数据湖进行数据开发，实现数据开发流程高度透明化、可配置化、大大提高数据类应用维护、核对、优化、改造等工作的效率和便捷性

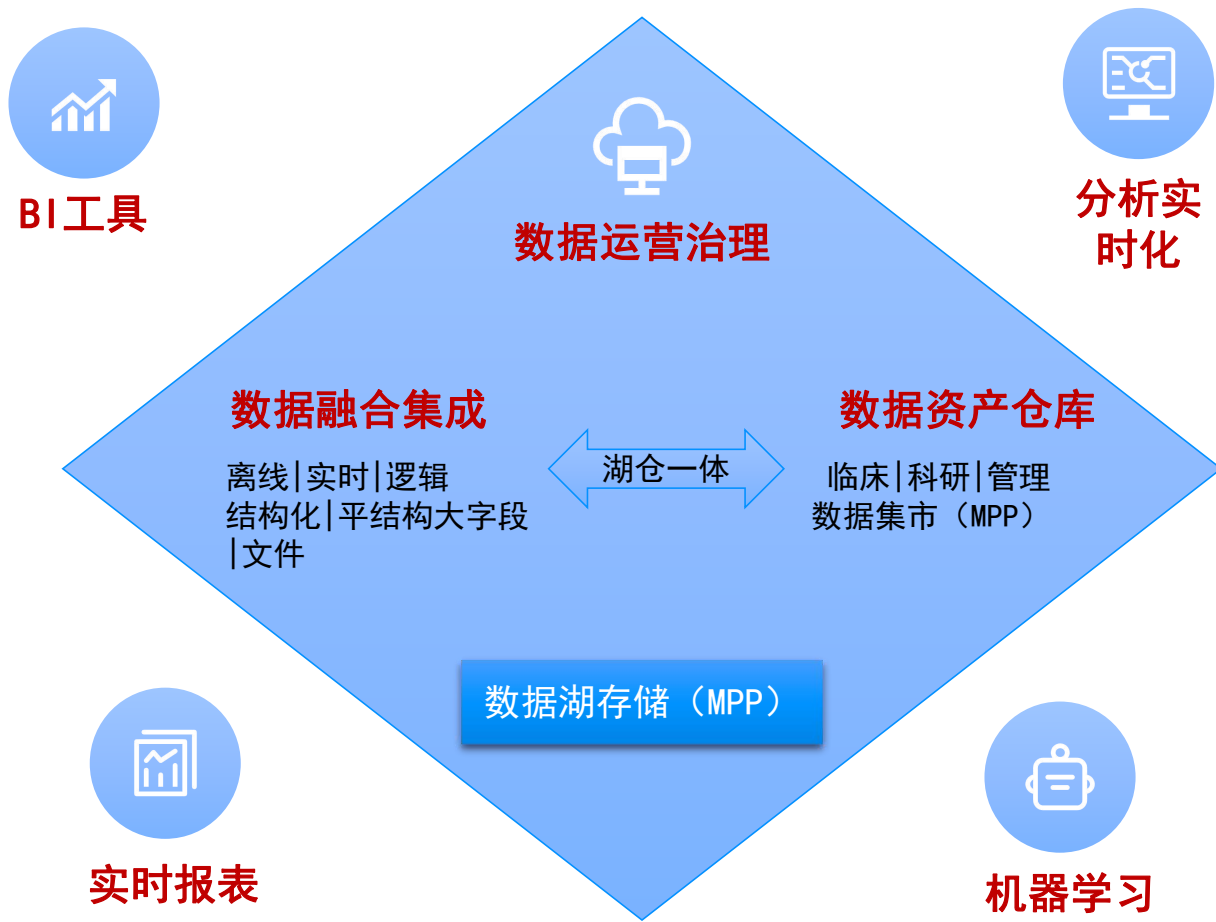
基于标准数据层的数据供给：

- 目标应用建设方在标准数据层或应用数据层的基础上，使用数据湖工具制作API，或通过建立同步作业将所需数据写入应用侧的数据库，供智慧屏应用使用或直接调用，实现数据展示





数据湖湖仓一体，提升分析效率



统一分析、减少迁移

- 湖仓协同，统一分析、2天到1小时，效率提升50倍
- 湖内建仓，数据不出湖，一套架构，减少80%数据迁移工作量

秒级入湖，毫秒分析

- 数据入湖，从T+1到T+0
- 在线分析，毫秒响应

数据资产，行业沉淀

- 统一管理元数据，实现数据资产盘点
- 医疗行业经验沉淀，国考指标模板沉淀

没有数据和湖时，逐个给业务库写统计分析脚本，然后人员汇总，至少两天。有湖仓时，一个查询引擎搞定



PART 04

平台功能展示

Platform function display

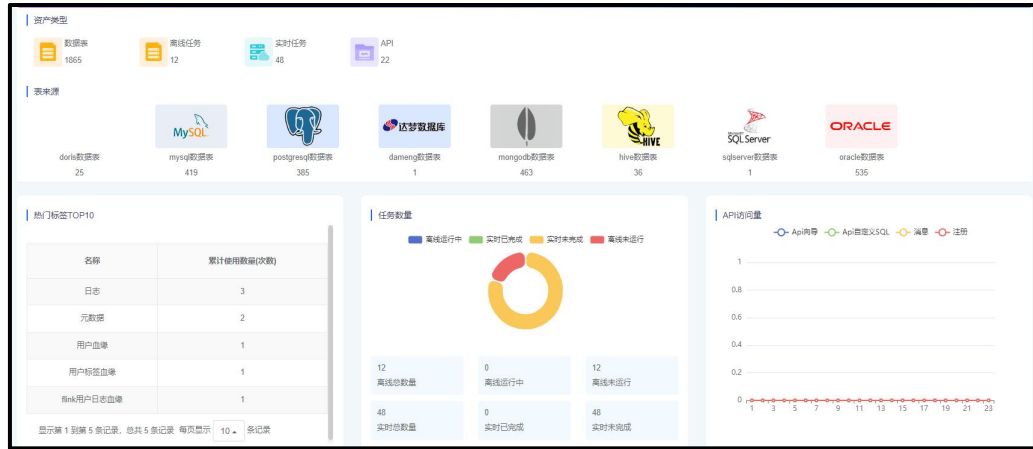
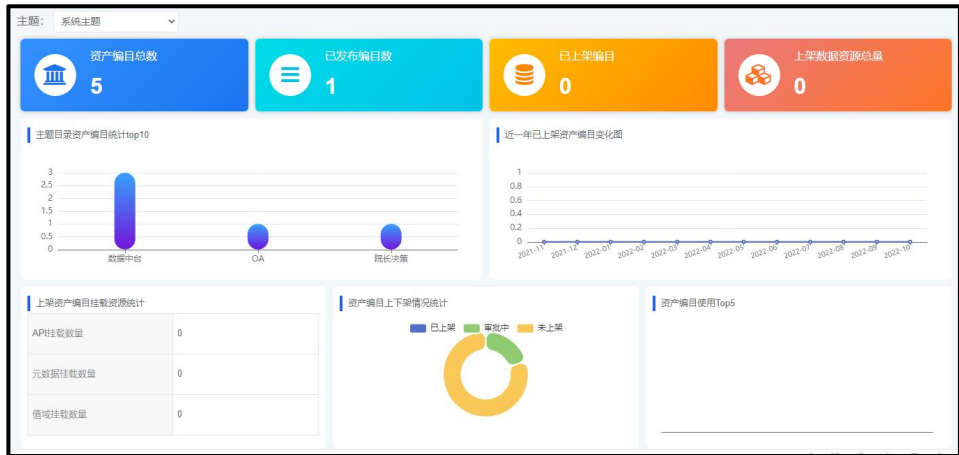




数据资产平台

企业数据资产门户，全域数据资产地图，在识别出自身数据资产的基础上，进一步构建数据资产目录，帮助用户更好的理解、使用以及分析数据。

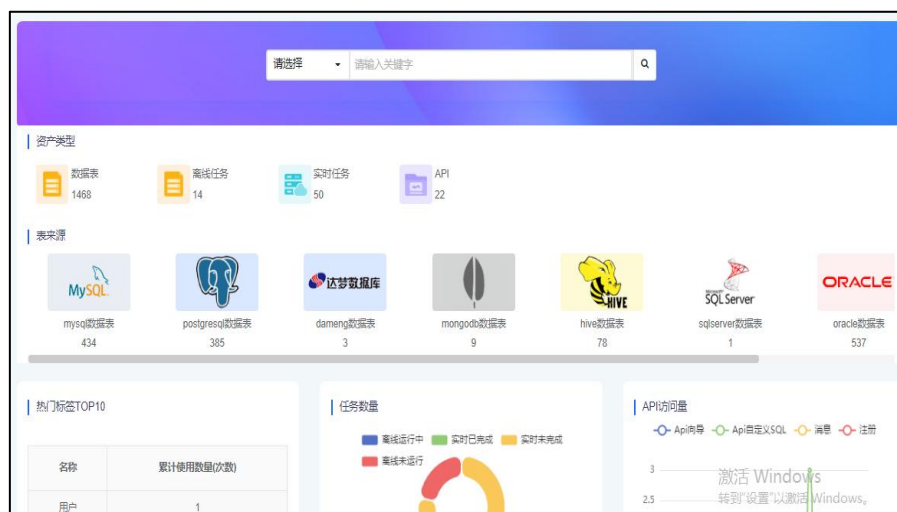
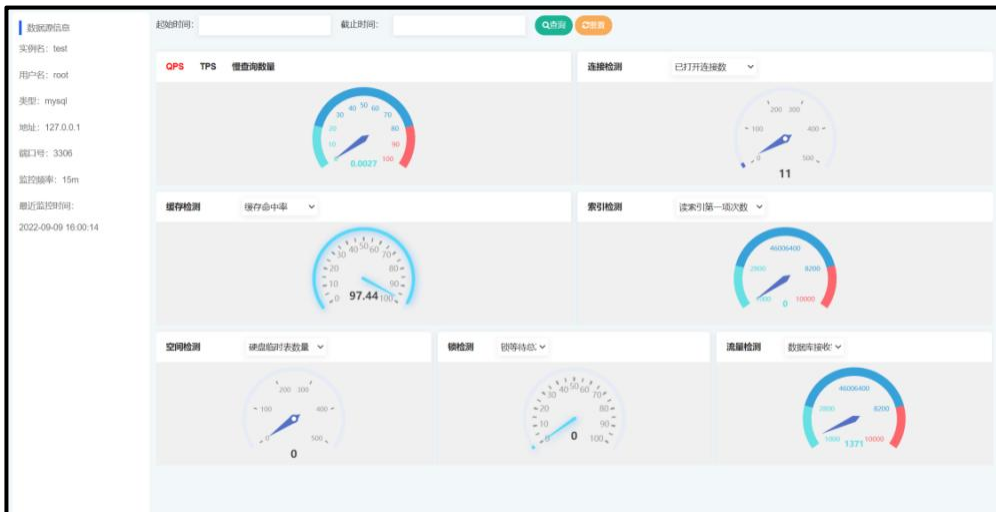
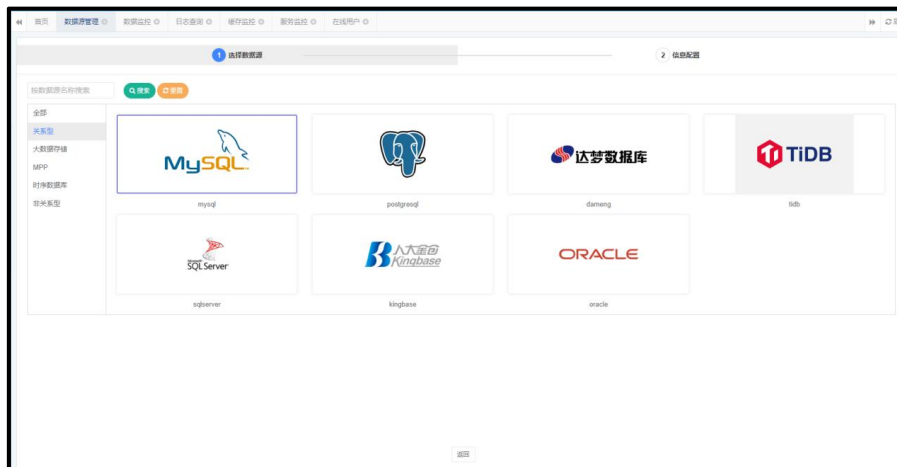
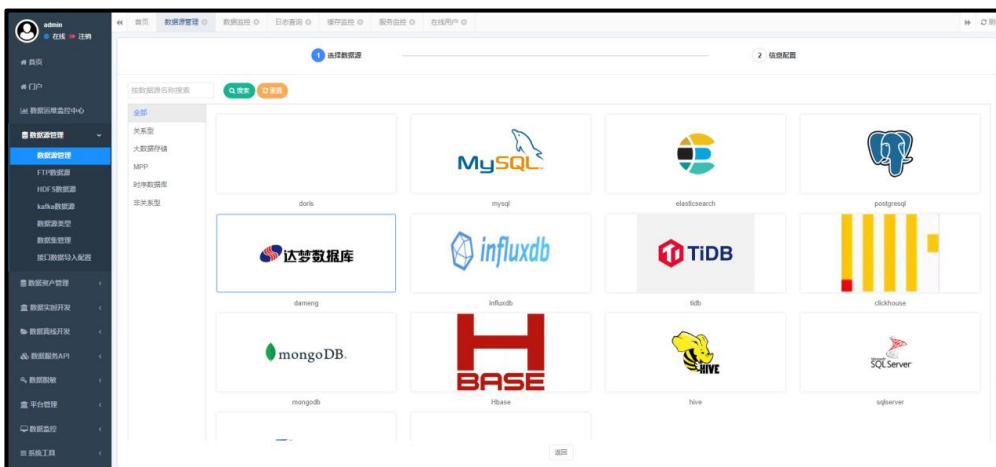
数据资产主要包含元数据采集、元数据搜索、数据血缘、数据模型、数据资产编目、资源上下架、资产市场、资源申请审批等能力。





数据资产平台：数据源管理

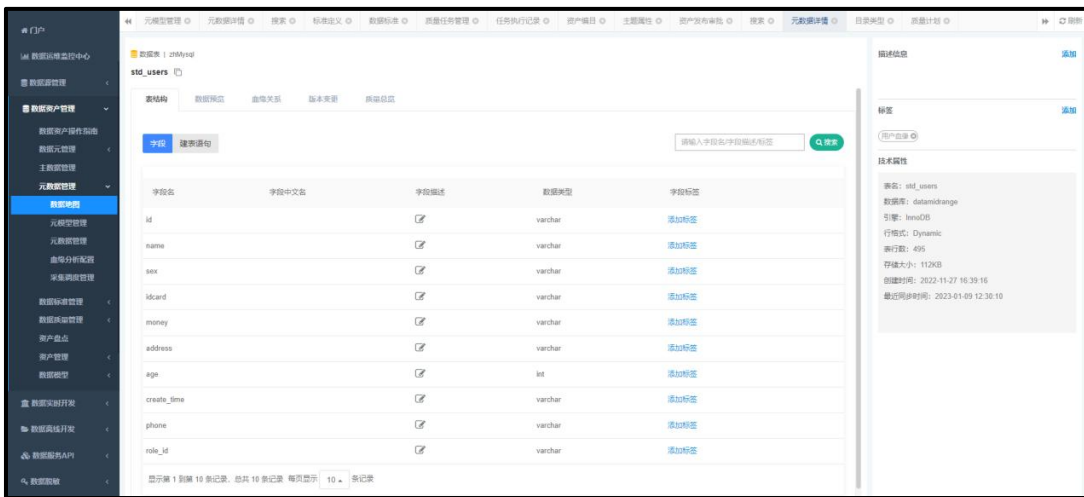
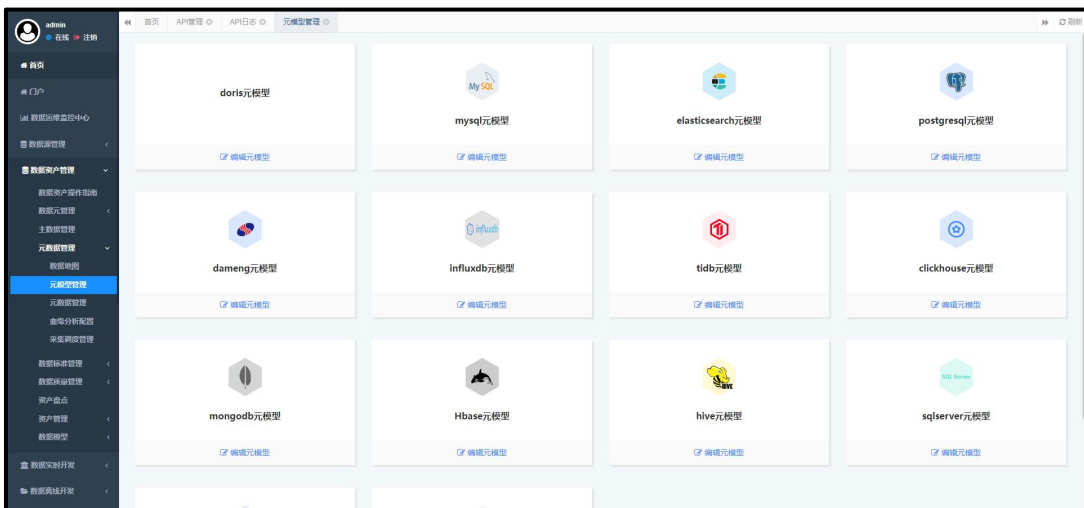
现在平台中支持的数据源类型除了包括mysql oracle 等常用关系数据源外，还支持达梦，金仓等国产化数据库，以及hbase、hive、doris等大数据存储类型。

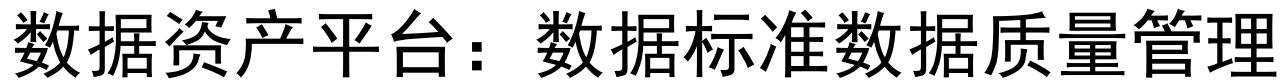




数据资产平台：元数据管理

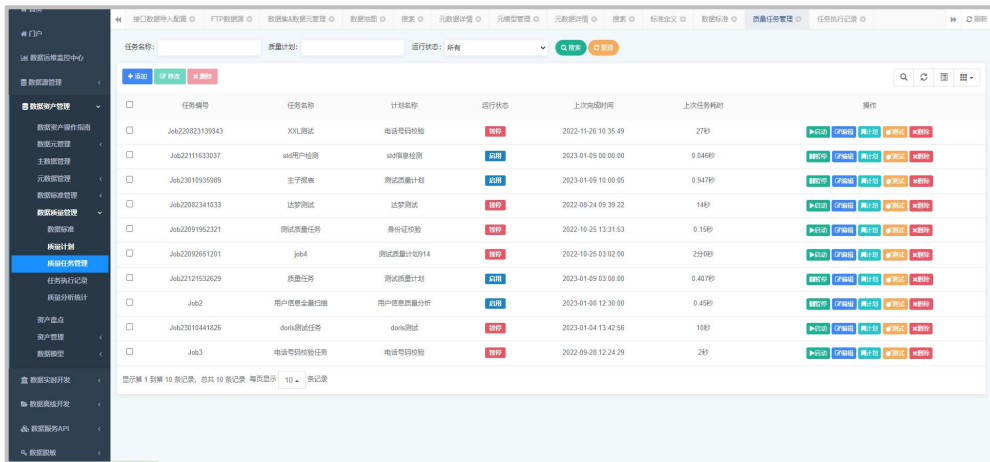
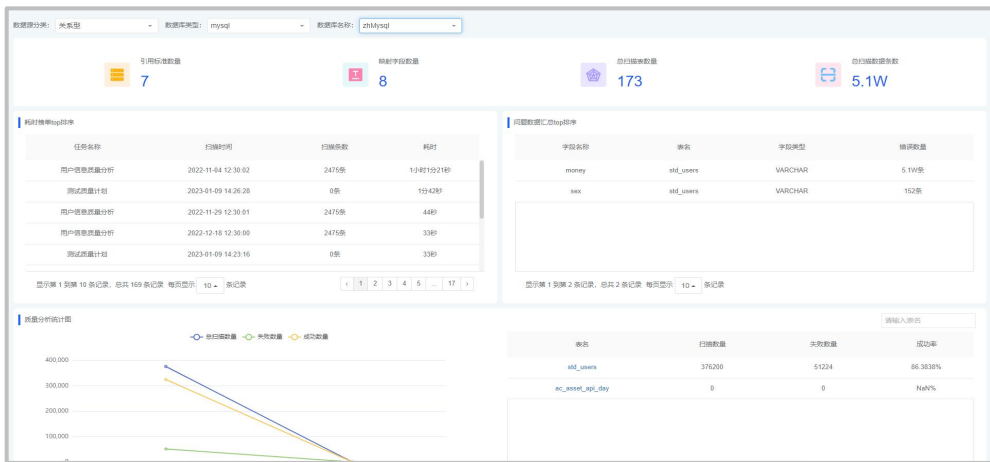
元数据是用来描述数据的数据，元数据管理功能可以查看技术元数据、业务元数据、管理元数据，帮助各行各业用户获得更多的数据洞察力，进而挖掘出隐藏在资源中的价值。通过元数据管理可以快速理清数据资源，了解数据来龙去脉，构建数据地图。





以数据标准为数据检核依据，数据质量可以监控到数据开发全流程中涉及的各个数据实体，实现数据质量问题的全面筛选和精准定位，

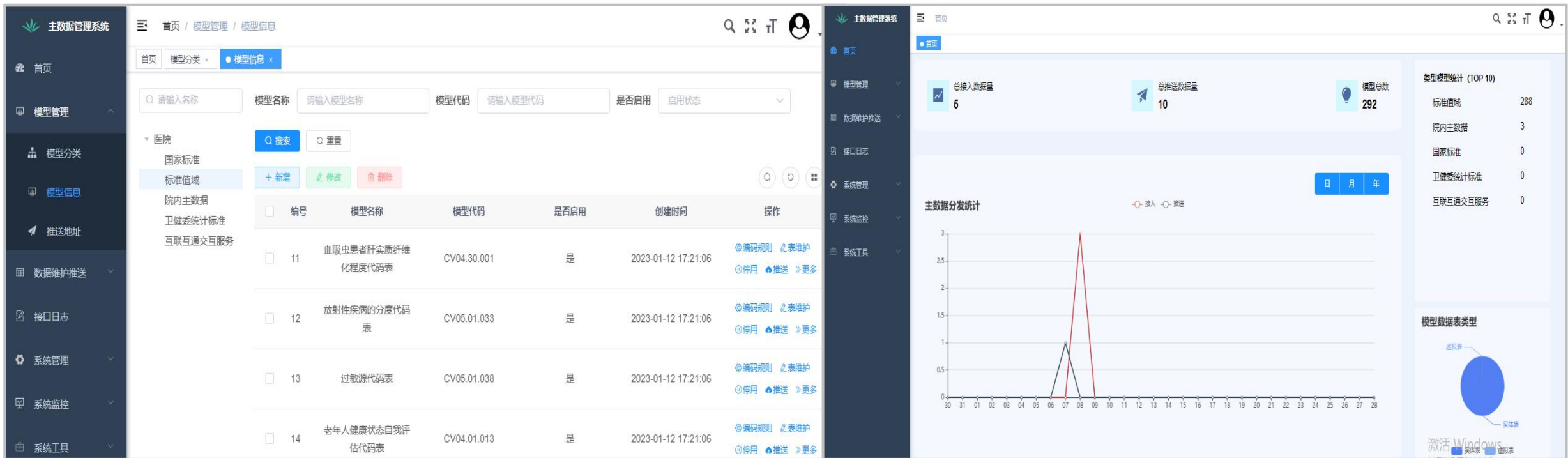
以可视化方式实现对数据质量的全局掌控和管理。数据质量主要包含数据标准管理、质量监控以及质量报告等能力。





数据资产平台：主数据管理

主数据管理作为数据治理体系的重要组成部分，保证各系统间共享数据的正确性、通用性、可控性、一致性、完整性，加速推进数字化转型的平台级产品。主数据不是企业内所有的业务数据，只是有必要在各个系统间共享交互的数据才是主数据，比如大部分的交易数据、帐单数据等都不是主数据，而像描述核心业务实体的数据，而像人员、科室、帐户、组织单位、员工、合作伙伴、药品等都是主数据。



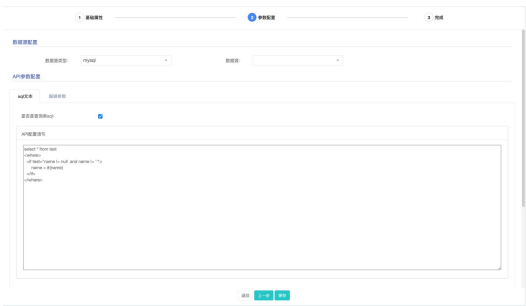


数据服务平台-API低代码开发

可视化“零代码”生成API，3分钟让数据快速服务化，提供开箱即用的API服务能力和轻量级ESB服务总线，精细化的数据安全管控，提高数据开放与共享效率。

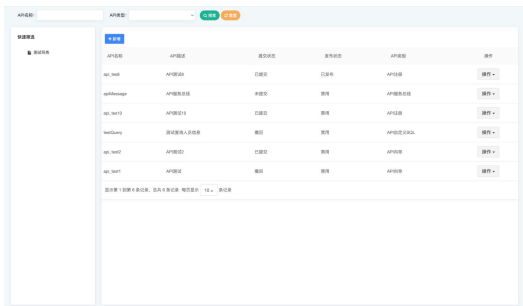
API创建

- 支持20多种数据库创建API
- 支持单表和夺标创建API
- 支持其他平台API注册
- 支持API服务编排式创建



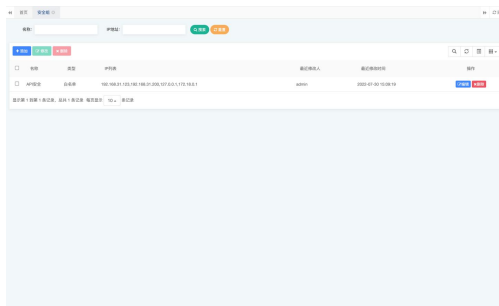
API管理

- API告警配置管理
- API标签目录分类管理
- 支持API调用限流管理
- 支持API停用和启用



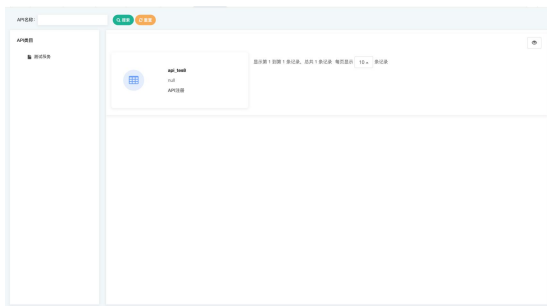
API安全

- API日志审计管理
- 支持API安全组访问
- 支持API行级权限配置
- 支持2种API加密方式



API市场

- 支持多重角色视角查看
- 支持API市场搜索查询
- 可形成企业/部门级API市场
- 支持API导入导出



数据源挂载



API编排



网关测试



API市场



API审批

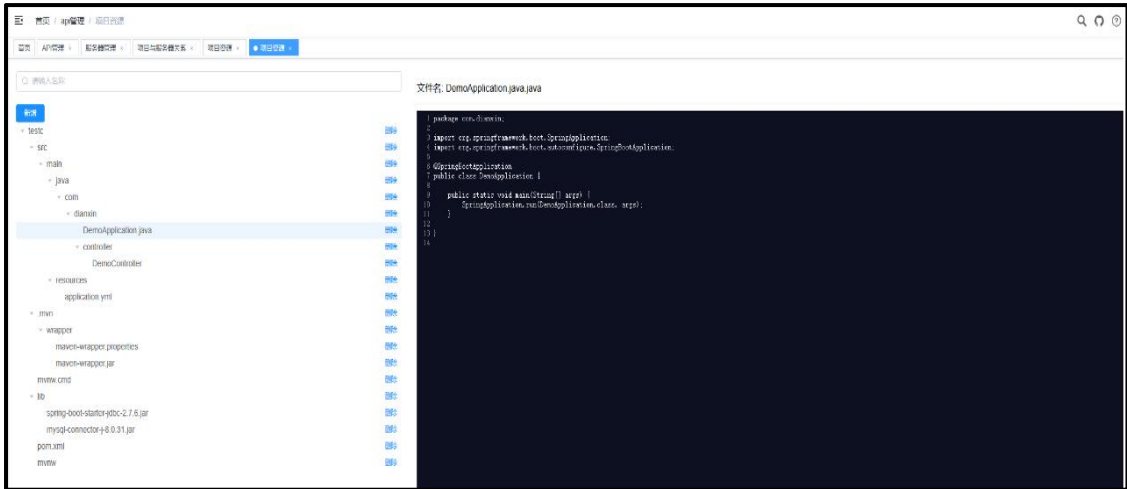
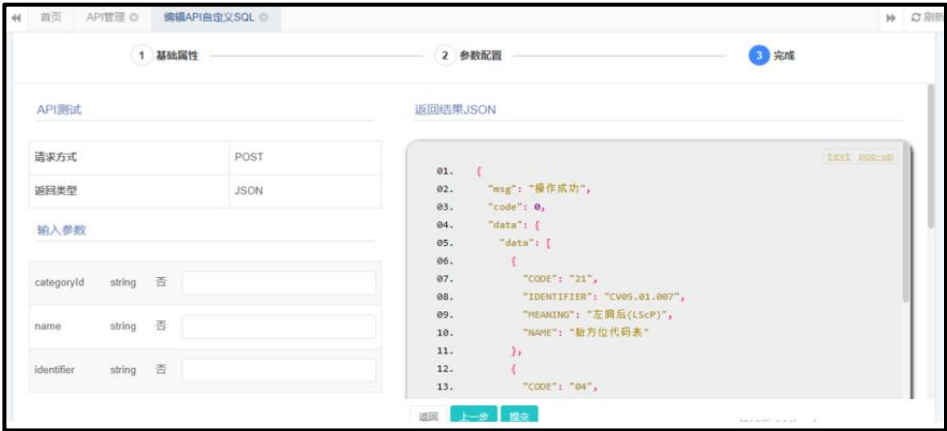


服务共享



数据服务平台

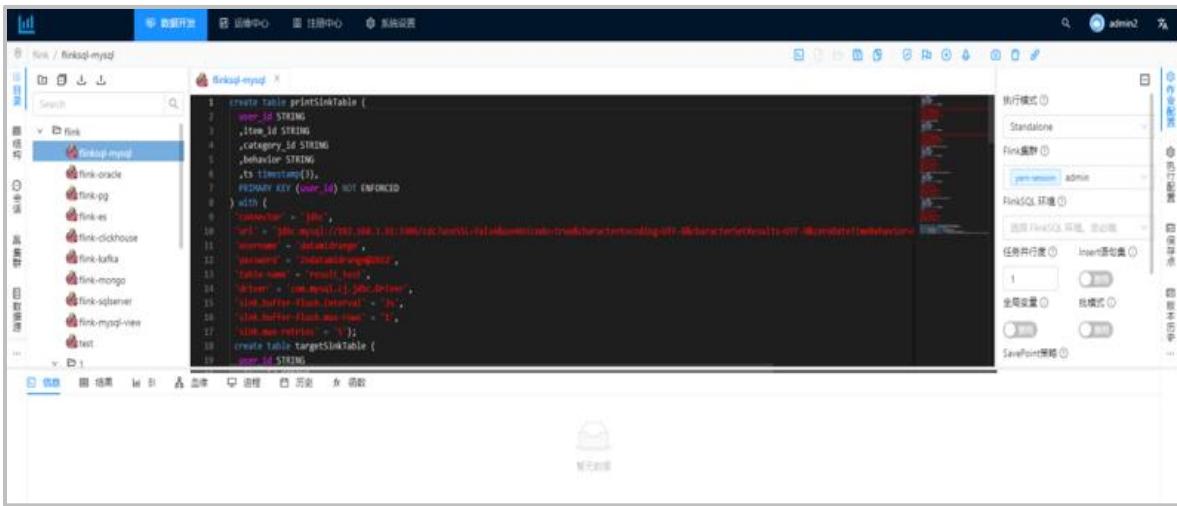
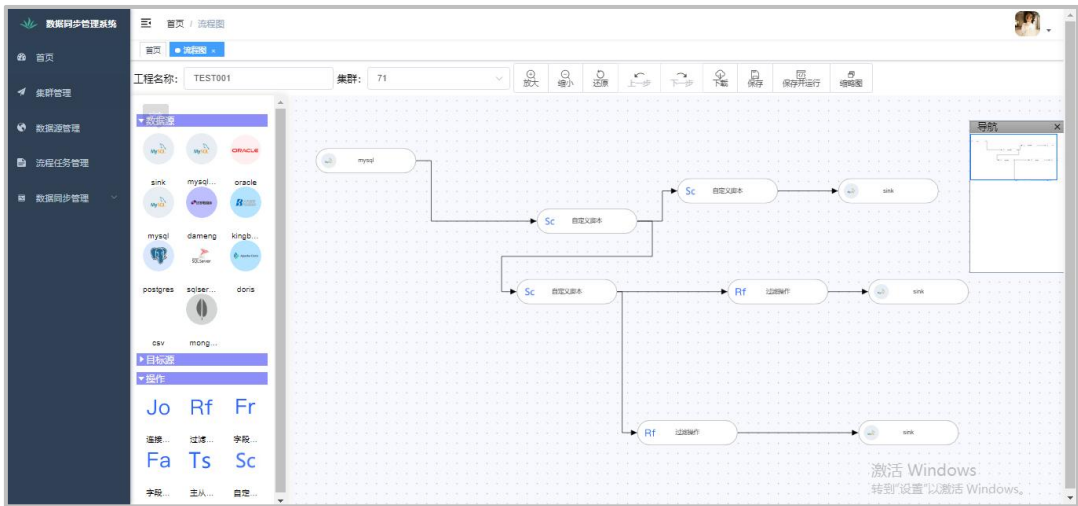
平台提供六种API注册类型，满足API注册端的多样性。通过API接口的快速开发平台，以低代码的方式，通过配置实现API接口的快速开发部署。支持 API 服务的全生命周期管理，覆盖 API 创建、第三方 API 注册、发布、审批、下线、历史版本管理的整个生命周期。同时追溯服务从生产到下线再到版本管理的各个环节。





数据开发平台

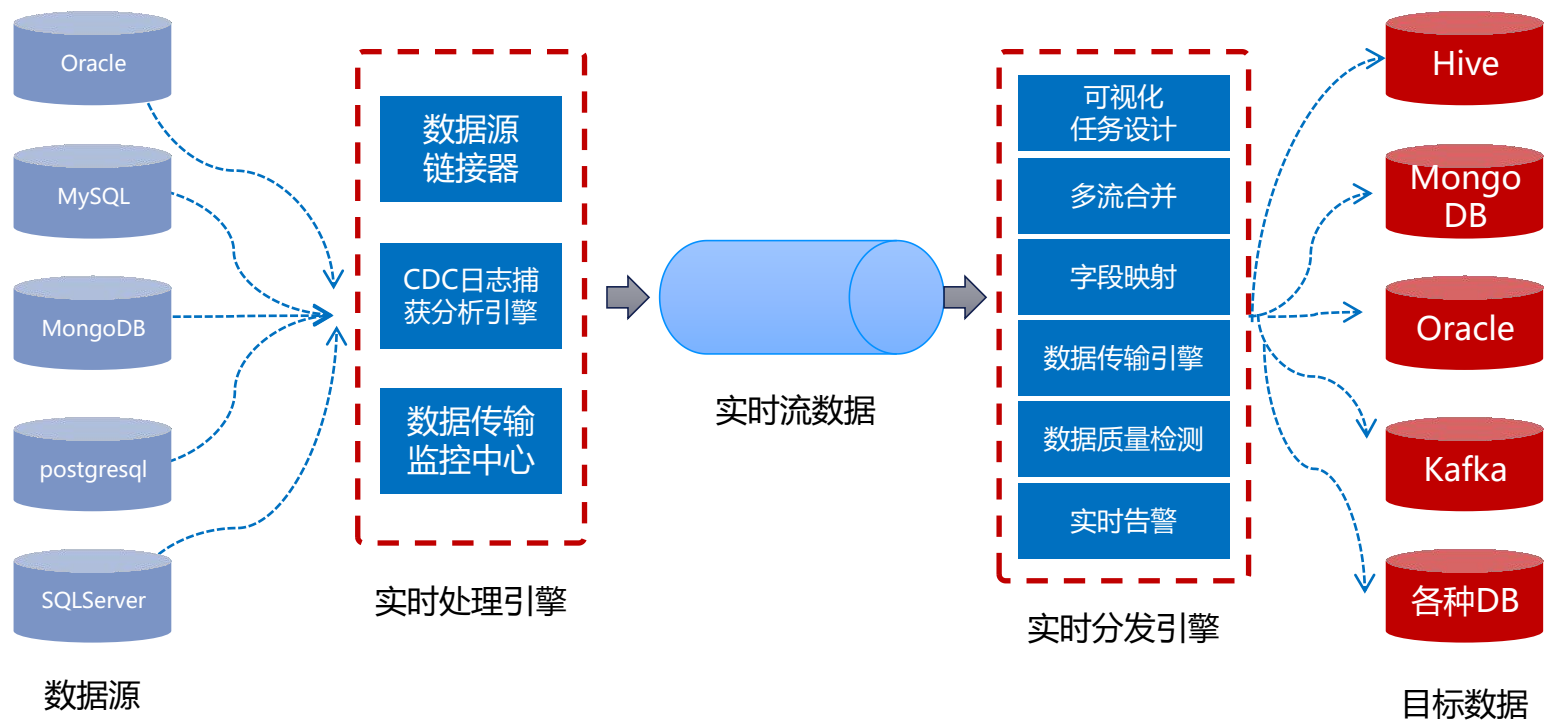
数据开发平台是一个数据实时同步+实时数据处理的数据平台，提供低代码形式的开发模式，用于企业多源数据融合，解决关系型数据库、非关系型数据库及大数据平台等异构数据源之间，大数据量高并发下的交互和数据同步问题，全面支持主流的开源数据库及商业数据库以及消息类中间件，非侵入式数据同步，基于数据库日志文件，无需安装插件，秒级同步DML和DDL操作，支持同构/异构数据源之间批量、增量、实时的数据迁移服务；应支持整库数据迁移，快速实现大量库表全量、增量数据同步。





数据开发平台

- 实时数据处理、集成、分发
- CDC数据变更捕获技术
- 增量数据实时同步、处理
- 变更数据实时发送
- 支持与ETL平台一站式集成



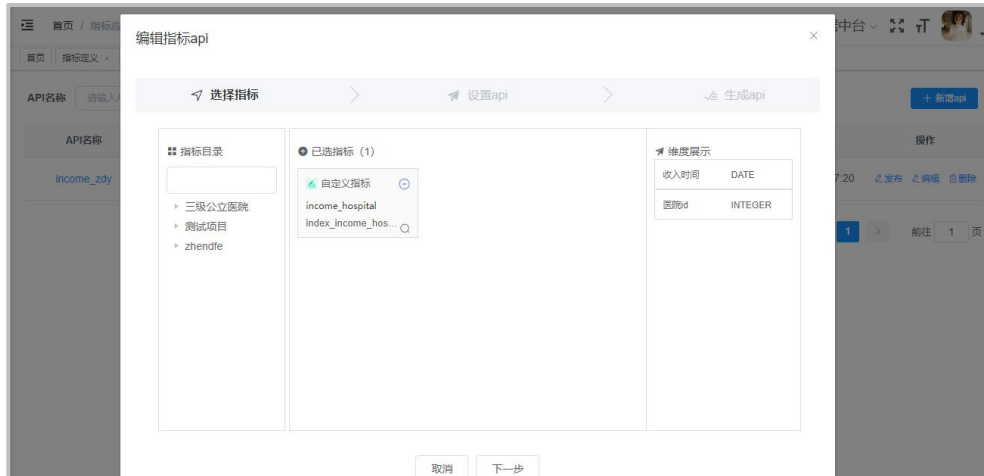
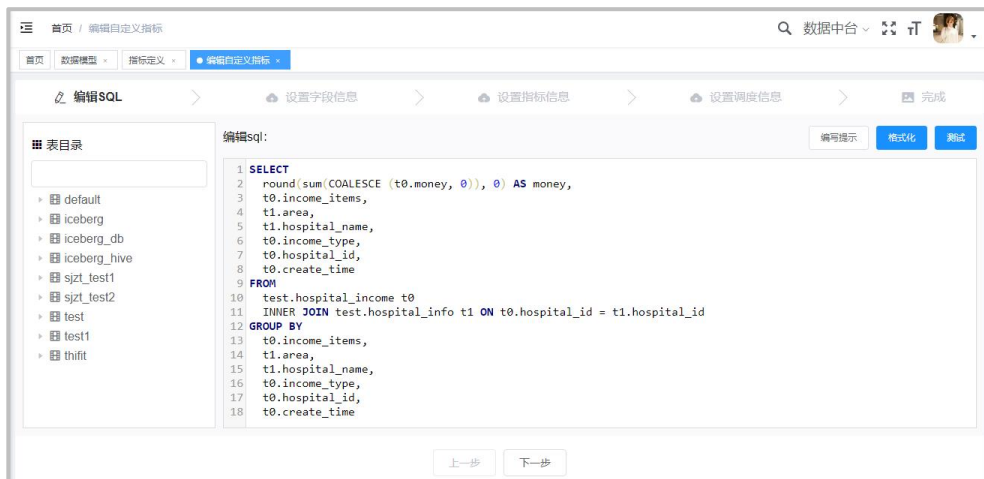
■ 实时数据传输——秒级实时数据传输能力



指标管理平台

一站式企业数据指标管理平台，涵盖指标定义、指标建模、指标可视化分析，实现指标全流程、一体化管理的系统。

旨在帮助各类企业建设指标体系，疏通指标数据，落地指标管理，发挥指标最大价值。落地指标数据结果，消除数据二义性，实现指标数据的可视、可用、可管。

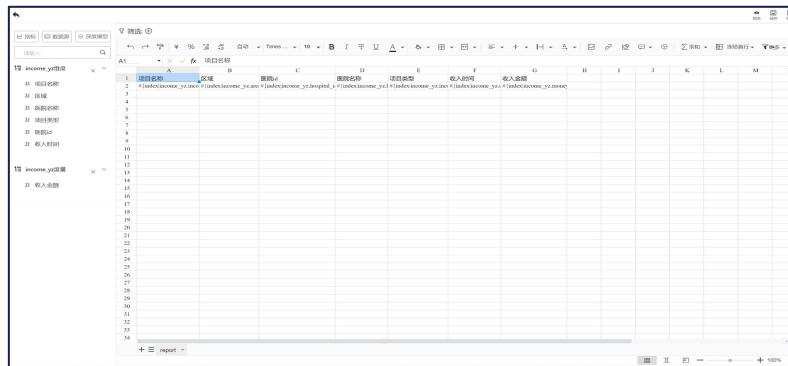




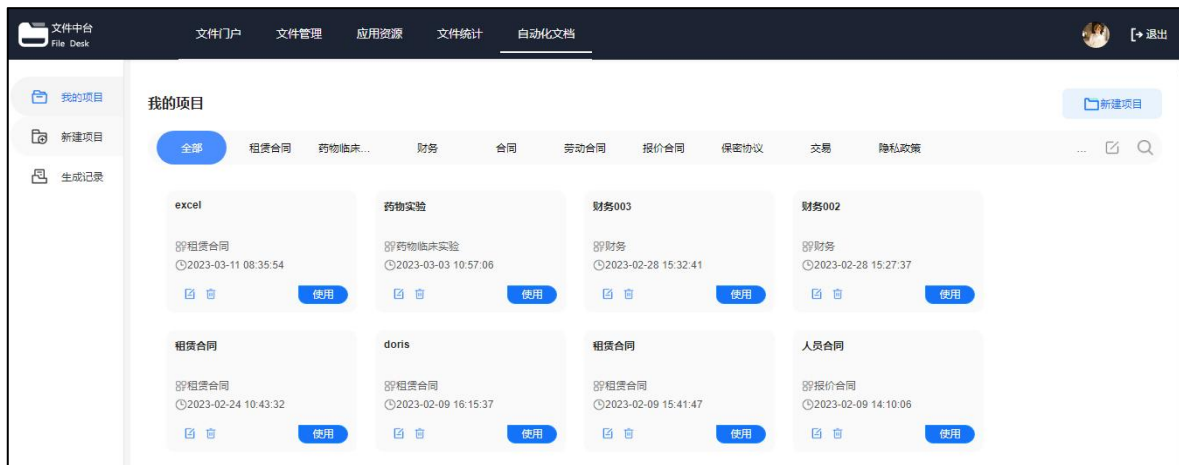
可视化数据分析平台

可以提供统一的自助式数据分析和决策辅助功能，支持强大的分析统计、数据挖掘功能，将不同的数据根据灵活的维度和度量进行多角度分析，将数据拖拽就可完成分析，深度挖掘数据价值。

将之前单一的数据进行多维度的关联分析，随时发挥数据创意，可以非常灵活的与数据交互，探索数据背后的原因并发掘更多价值，为决策制定提供有效的数据支撑。



文件中枢



The screenshot shows the 'File Desk' interface with a sidebar on the left containing navigation options like '药物实验' (Drug Experiment) and '类型: word' (Type: word). The main area displays a document preview for a 'word' file. The document content includes sections for '1. 知情同意书签署日期' (Informed Consent Signature Date), '2. 人口统计学资料' (Demographic Information), and '3. 病史' (Medical History). The demographic information is presented in a table with columns for '姓名' (Name), '筛选编号' (Screening Number), '性别' (Gender), '民族' (Ethnicity), '出生日期' (Date of Birth), '职业' (Occupation), '籍贯' (Place of Origin), and '单位' (Unit). The medical history section includes '家族病史' (Family History) and '个人病史' (Personal History).

➤ 实现企业和组织非结构化数据的统一存储、统一管理和统一服务。打通各业务系统，提供强大的文件中台管理能力。横向打通了非结构化数据孤岛，构建起统一的非结构化数据中台，实现数据互通、业务融合。

➤ 通过平台提供的无需写代码的文档自动化系统，可以将复杂繁琐的文档制作工作转化为高效的自动化应用，只要您会用WORD写文档，就会创建文档自动化工具。

实时湖仓一体化数据开发平台

感谢观看聆听

Your content is entered here, or by copying your text,
select Paste in this box and choose to retain only text.

