数据治理中心
数脉平台用户手册 V2.0 · 数据治理中心

ETL 调度中心
数据同步
数据加工
外部数据源
元数据管理
AI 同步生成器
AI 加工设计器
AI 全链路编排
AI 数仓分层建模器

ETL 调度中心

功能定位
ETL 调度中心是数据治理流水线的 “时间与依赖控制器 ”。它把数据同步、清洗加工、 dbt 建模、质量校验和元数据刷新组织成有先后依赖的工作流，按日、周或手工触发，并保留每次运行的实例、日
志和重试记录。数脉通过统一门户和安全免登录票据打开 DolphinScheduler
，用户不需要再输入一套调度平台账号。

图 7-1 DolphinScheduler 集成诊断、版本与安全跳转入口

适用角色

ETL
管理员、数据平台运维人员、数据开发人员和值班人员。业务用户通常不直接进入调度中心，而是在大屏或报
表中消费调度产出。

页面入口
数据治理中心 → ETL 调度中心（ /bi/etl/ds）。页面内嵌 DolphinScheduler 3.4.3。

前置条件
DolphinScheduler 服务健康，调度数据库和 Worker 正常。
当前账号已开通 ETL 调度中心访问权限，并完成平台账号与调度租户的映射。
后端已配置免登录交换地址和安全票据密钥。
目标工作流所需的 Doris、SeaTunnel、 Hop、dbt 等执行环境可用。

操作步骤
1. 从菜单进入后，页面先显示 “集成状态提示 ”
。绿色表示已连接，黄色或红色表示配置关闭、账号未映射或服务异常。
2. 状态正常时，系统签发短时、一次性的访问票据并自动打开 DolphinScheduler。
3. 在工作流定义中按项目查看已有 DAG；不要把生产工作流直接改成临时试验版本。
4. 打开 “工作流实例 ”查看运行状态、开始结束时间、耗时和失败节点。
5. 失败时先看任务日志定位原因；修复数据、依赖或环境后，再执行 “重跑 ”或 “补数 ”。
6. 需要临时验证时使用手工运行，并在运行说明中写明业务原因、影响范围和期望完成时间。
7. 完成后回到数脉的元数据、质量或发布页面核对产出，不要仅以 “调度成功 ”判断数据正确。

输入与输出

项目 说明

输入 工作流定义、调度周期、依赖关系、运行参数、补数日期

输出 工作流实例、任务实例、运行日志、产出表和状态回写

失败处理 定位失败节点后重试；涉及重复写入时必须依赖幂等键或批次号

宏澜家电示例
每天 02:00 启动 “销售订单全链路 ”：先同步 ERP 订单到 ODS，再清洗为 DWD，汇总区域与品类 DWS，执行 dbt 质量测试，刷新元数据，最后登记 Superset 数据集刷新。若 02:10 发现 ERP
连接超时，值班人员先确认源库恢复，再对同步节点补跑；下游任务依赖未满足时不会提前执行。

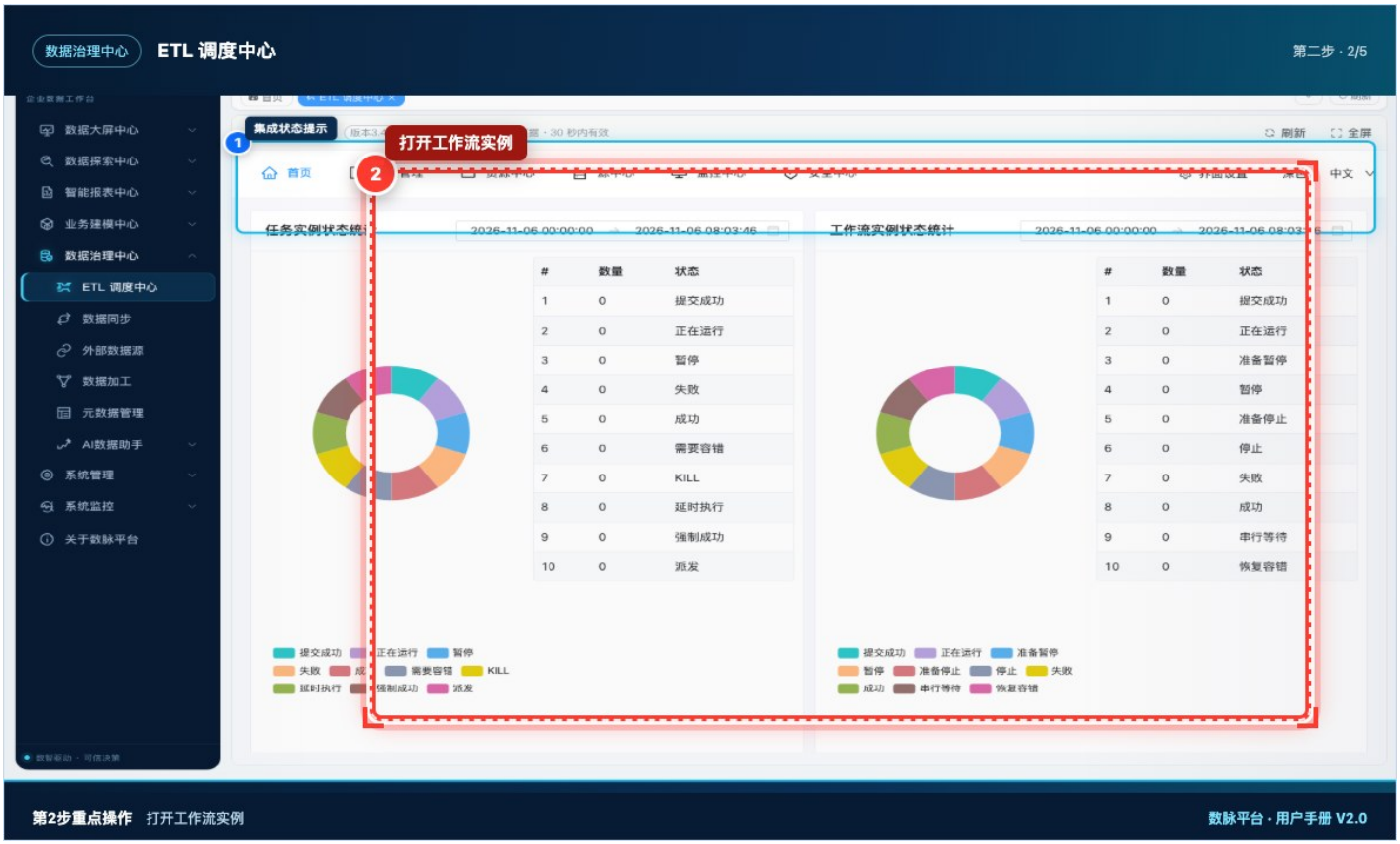
客户应用演示：检查每日销售同步工作

流
业务情境： 值班人员早上发现经营看板未刷新，需要检查 02:00 的调度实例。

第一步：进入 “数据治理中心 → ETL 调度中心 ”

步骤图 35-01 第1步：进入 “数据治理中心 → ETL 调度中心 ”。红框标出进入数据治理中心 → ETL 调度中心。
进入 “数据治理中心 → ETL 调度中心 ”，查看集成状态和诊断信息。

第二步：打开 “工作流实例 ”



步骤图 35-02 第2步：打开 “工作流实例 ”。红框标出打开工作流实例。
打开 “工作流实例 ”，定位当天销售全链路实例。

第三步：查看开始时间、结束时间、耗时和失
败节点



步骤图 35-03 第3
步：查看开始时间、结束时间、耗时和失败节点。红框标出查看开始时间、结束时间、耗时和失败节点。

查看开始时间、结束时间、耗时和失败节点。

第四步：打开失败任务日志



步骤图 35-04 第4步：打开失败任务日志。红框标出打开失败任务日志。

打开失败任务日志，区分源库连接、权限、数据质量或下游依赖问题。

第五步：修复原因后按依赖顺序补跑



步骤图 35-05 第5步：修复原因后按依赖顺序补跑。红框标出修复原因后按依赖顺序补跑。

修复原因后按依赖顺序补跑，并回到看板核对数据更新时间。

最终结果与效果
能够定位失败节点和影响范围，补跑后看板恢复更新；历史和重试记录均可追溯。

操作提示： 当前演示环境可能显示 DISABLED “
，表示入口已实现但外部调度集成未启用，不是菜单损坏。

当前演示环境说明
截图中页面显示 “DolphinScheduler 集成已停用 ”，诊断码为 DISABLED 。这表示页面和免登录通道已经实现，但当前运行配置主动关闭了该集成，并非页面链接失效。启用方法由部
署管理员在环境配置中完成，普通用户无需修改浏览器地址。

常见问题
一直停在 “正在校验权限并签发安全访问票据 ”。
检查后端、调度服务和账号映射；票据通常只有几十秒有效，超时后点击 “重新检测并打开 ”。

提示账号未开通。 平台账号还没有映射到调度租户，联系 ETL 管理员开通。

调度成功但报表没有新数据。
依次检查产出表行数、批次时间、质量规则和数据集刷新，不要只重复点击重跑。

开发人员参考
页面： ui/src/views/bi/dolphinscheduler/index.vue
接口： GET /bi/dolphinscheduler/status、 GET /bi/dolphinscheduler/launch、 GET /bi/dolphinscheduler/exchange
权限： bi:etl:ds
门户只传递一次性票据，调度服务账号和密码不会发送到浏览器。
iframe 使用受限 sandbox、 referrerpolicy=no-referrer，并设置加载超时和错误诊断码。

数据同步

功能定位
数据同步把 ERP、业务库或文件等源数据按白名单配置抽取到 Doris 的 ODS 层。它解决 “ 数据如何稳定进入平台 ”的问题，重点不是写复杂 SQL
，而是把源表、目标表、全量或增量模式、批次、重试和运行记录管理清楚。

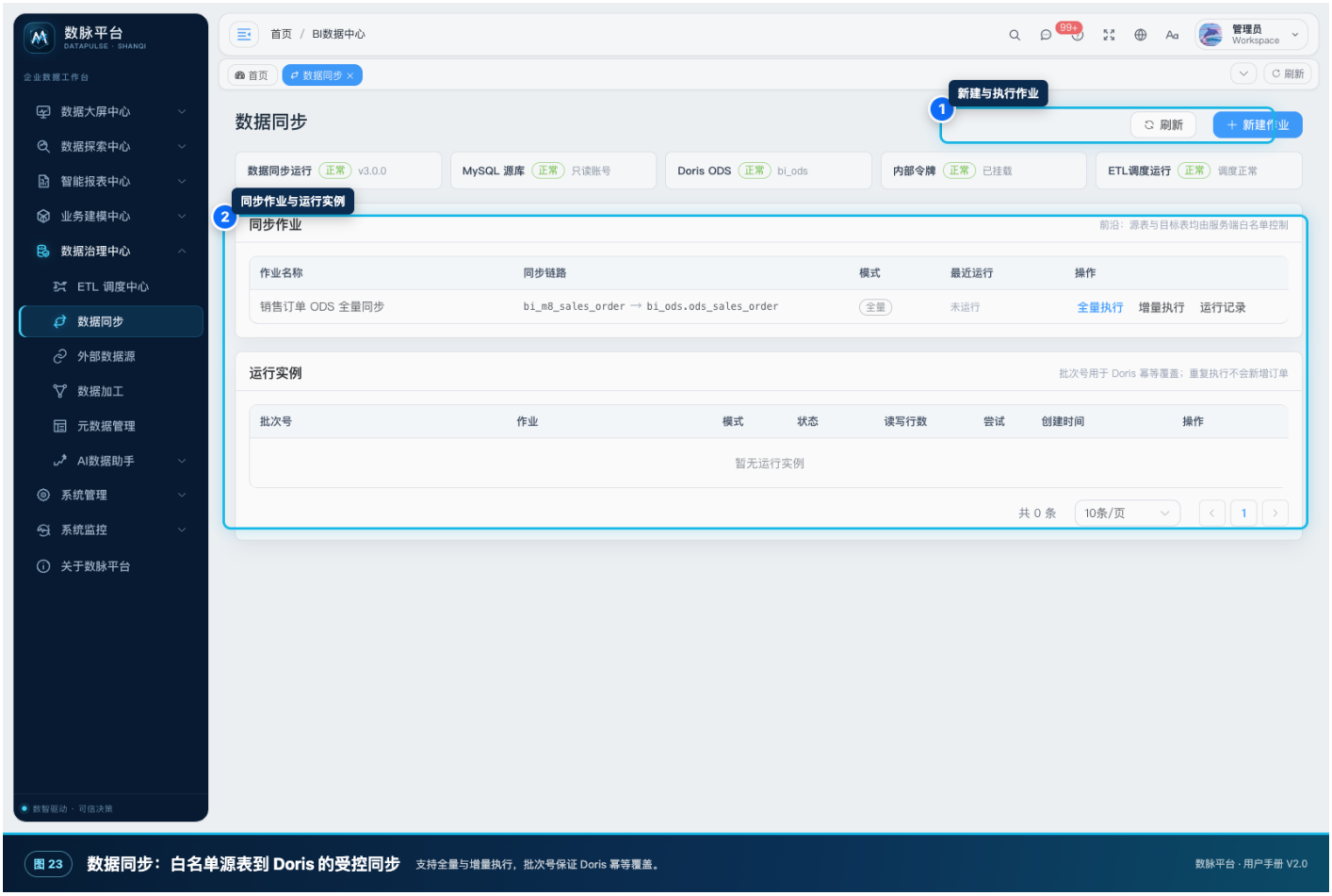


图 7-2 白名单源表到 Doris 的受控同步

适用角色
ETL 管理员、数据开发人员和数据运维人员。

页面入口

数据治理中心 → 数据同步（`/bi/etl/seatunnel`）。

页面结构与关键字段

区域	作用
集成状态	显示数据同步集成、MySQL 源库、Doris ODS、内部令牌和 ETL 调度是否可用
同步作业	展示作业名称、同步链路、模式、最近运行和操作
运行实例	展示批次号、作业、模式、状态、读写行数、尝试次数和创建时间
操作	全量执行、增量执行、运行记录、重试

操作步骤

1. 确认状态栏中的源库、Doris ODS 和调度服务均为可用；任何一项异常都先处理环境。
2. 点击“新建作业”，选择服务端白名单中的源表和目标 ODS 表。页面不允许手工输入任意库表。
3. 选择同步模式：首次接入或需要重建时使用全量；日常增量按时间戳、主键或批次水位同步。
4. 保存后先执行一次小范围验证，检查读取行数、写入行数和抽样数据。
5. 在“运行实例”中按批次号跟踪；重试应沿用原批次号，避免产生重复订单或重复库存。
6. 核对 Doris 目标表行数、主键唯一性和最新时间，再通知下游加工。

输入与输出

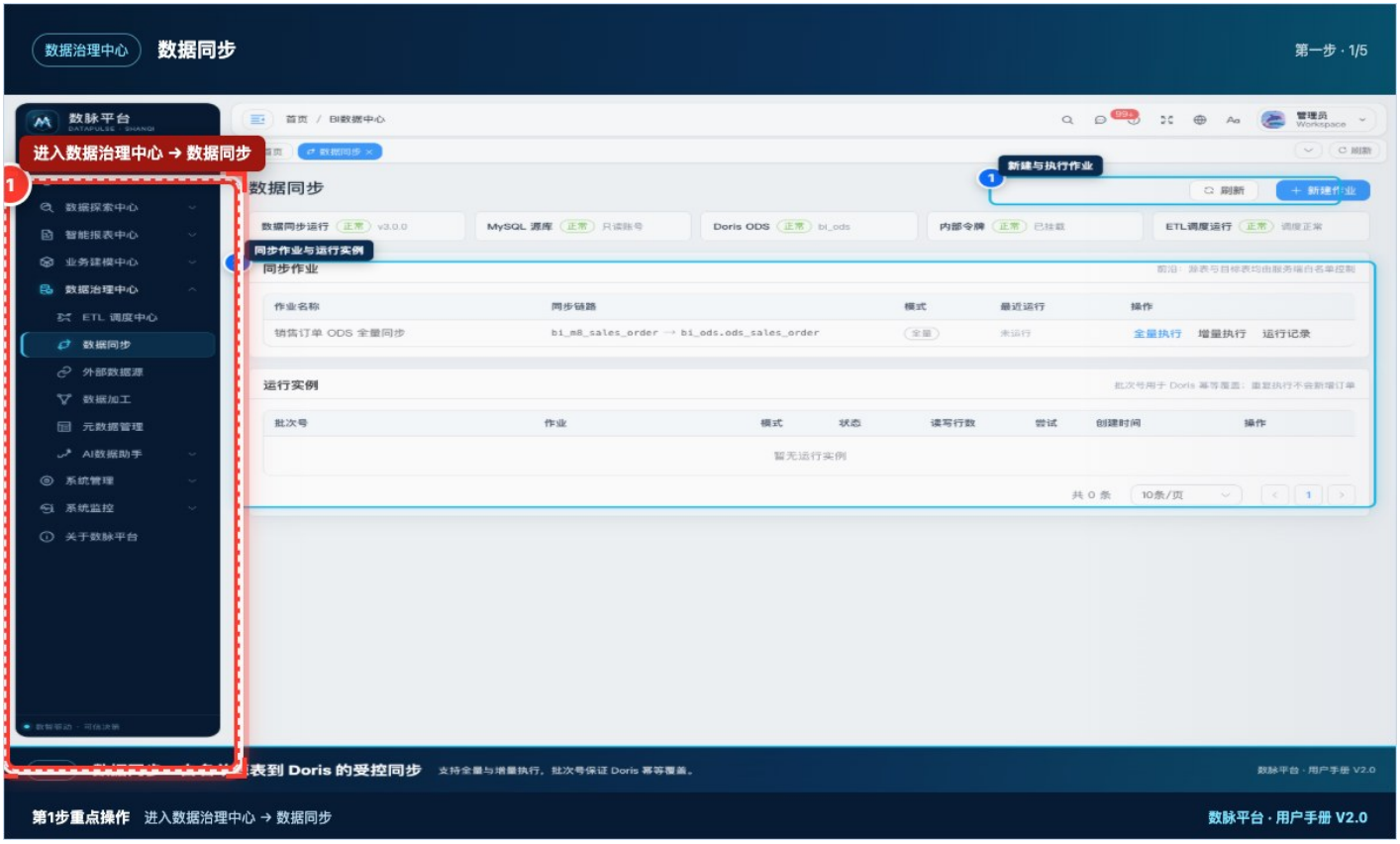
输入是源表、目标表、同步模式、字段映射和批次参数；输出是 ODS 表数据、批次记录、读写行数和运行状态。

宏澜家电示例

将 ERP 的 `sales_order` 和 `sales_order_detail` 同步到 `ods_sales_order` 与 `ods_sales_order_detail`。首次全量 18 万行，之后每小时按更新时间增量。某次网络中断后重试保持同一批次号，Doris 幂等覆盖，没有产生重复订单。

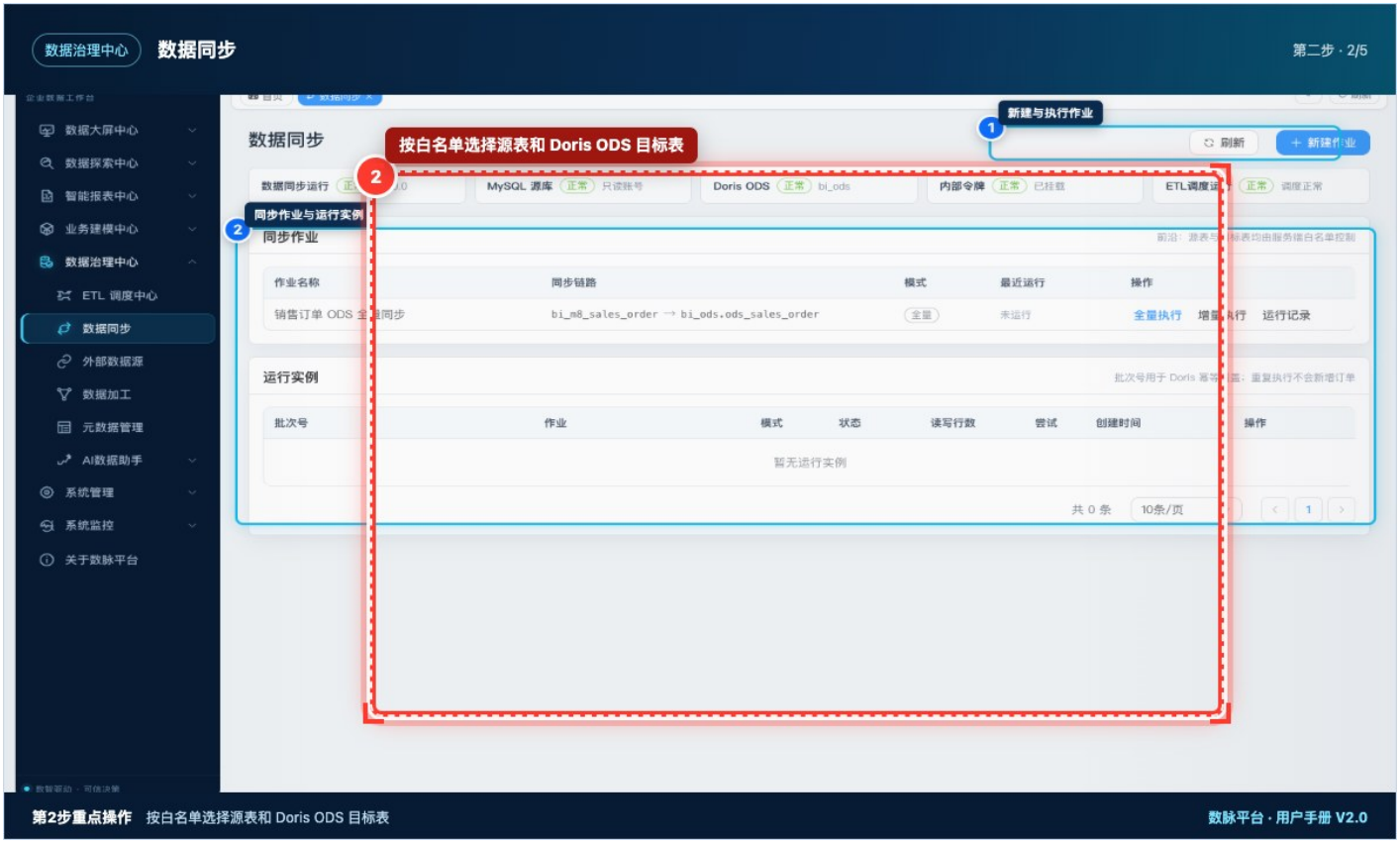
客户应用演示：把 ERP
销售订单同步到 ODS
业务情境： 需要首次同步宏澜 ERP 的销售订单、客户和产品表。

第一步：进入 “数据治理中心 → 数据同步 ”



步骤图 36-01 第1步：进入 “数据治理中心 → 数据同步 ”。红框标出进入数据治理中心 → 数据同步。进入 “数据治理中心 → 数据同步 ”，新建同步作业。

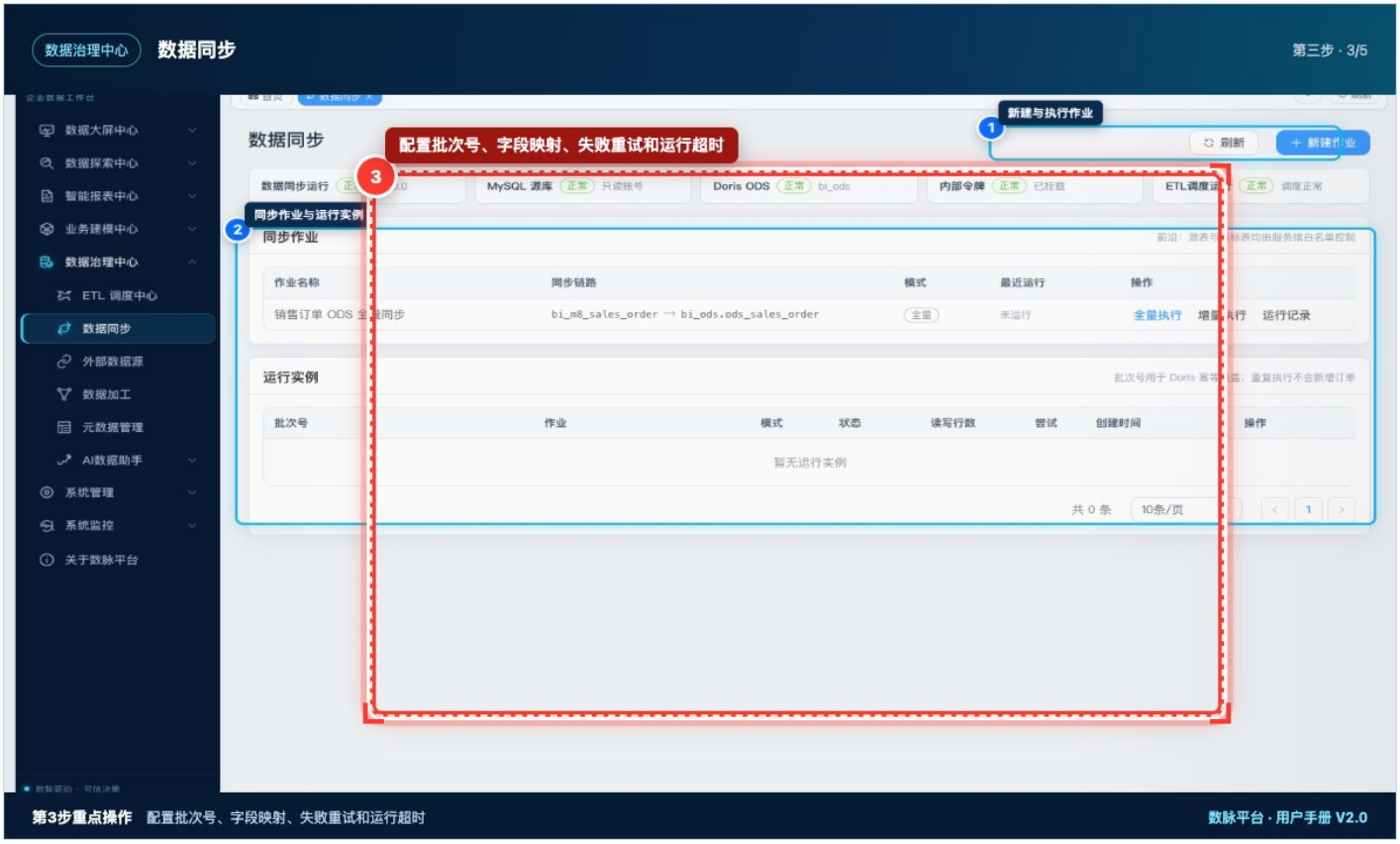
第二步：按白名单选择源表和 Doris ODS 目标表



步骤图 36-02 第2步：按白名单选择源表和 Doris ODS 目标表。红框标出按白名单选择源表和 Doris ODS 目标表。
按白名单选择源表和 Doris ODS 目标表，设置全量或增量模式。

第三步：配置批次号、字段映射、失败重试和

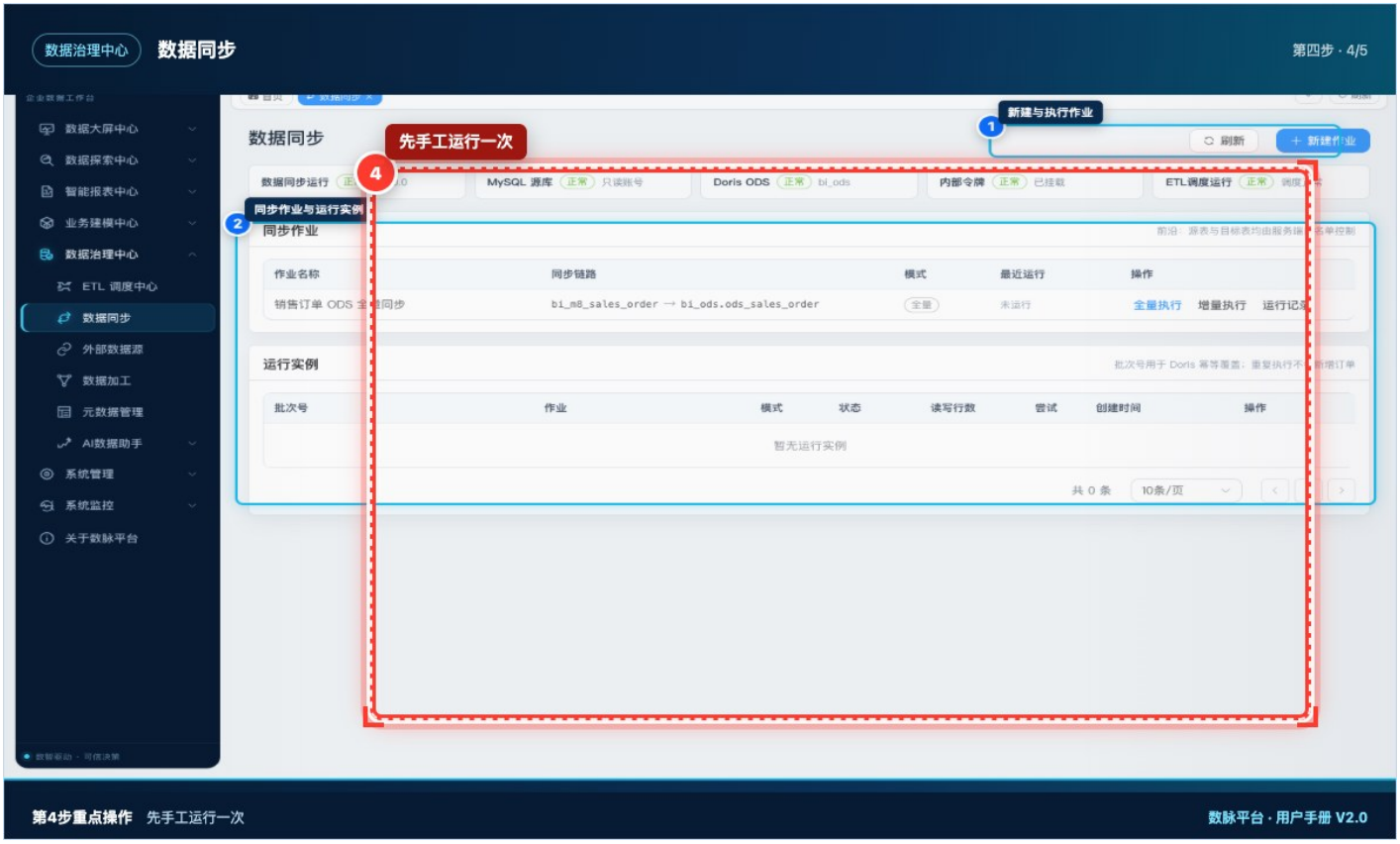
运行超时



步骤图 36-03 第3
步：配置批次号、字段映射、失败重试和运行超时。红框标出配置批次号、字段映射、失败重试和运行超时。

配置批次号、字段映射、失败重试和运行超时。

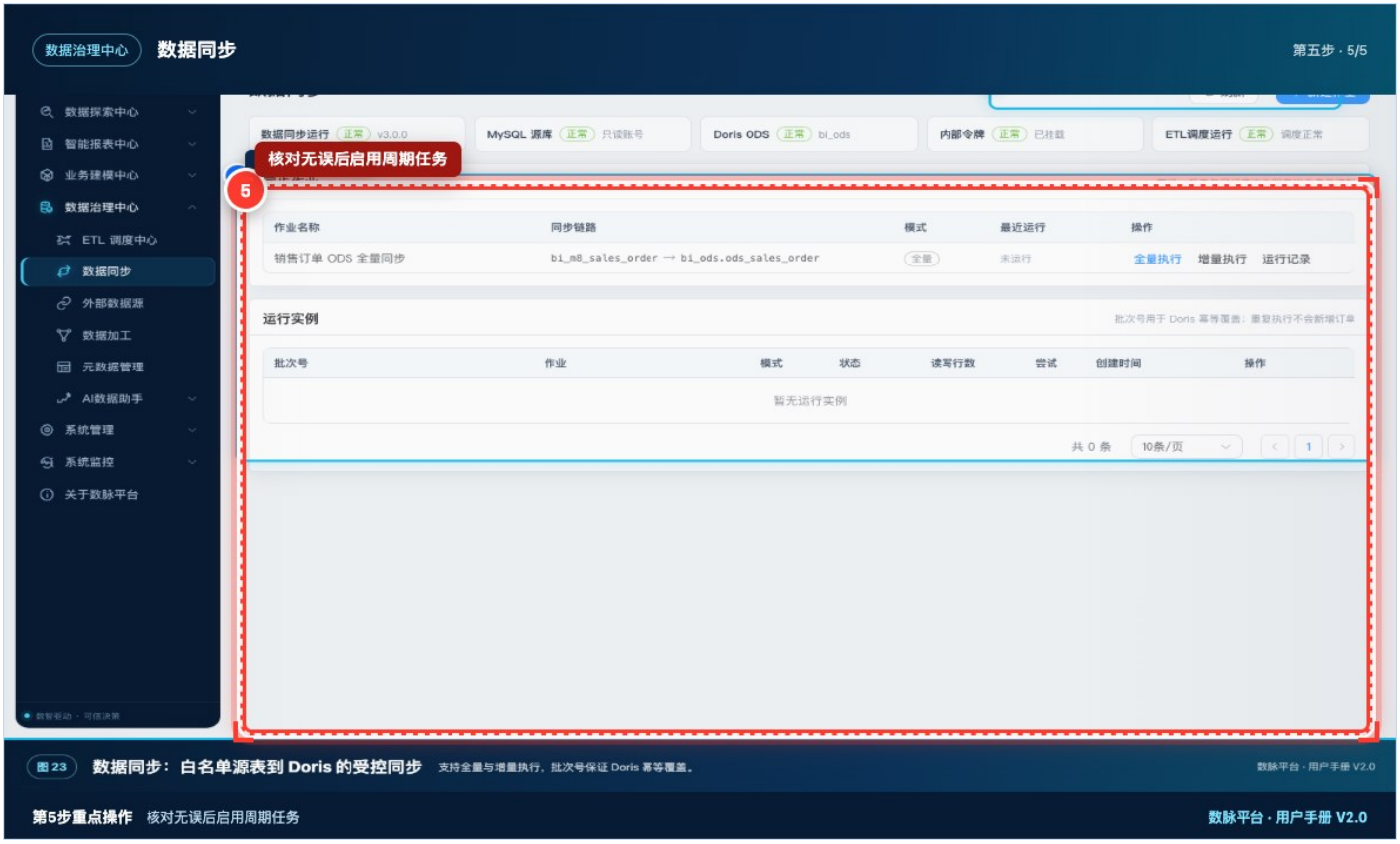
第四步：先手工运行一次


数据治理中心
数据同步
第四步 · 4/5
数据大屏中心
数据探索中心
智能报表中心
业务建模中心
数据治理中心
ETL 调度中心
数据同步
外部数据源
数据加工
元数据管理
AI数据助手
系统管理
系统监控
关于数脉平台
数据同步
新建与执行作业
刷新
新建作业
先手工运行一次
同步作业与运行实例
同步作业
作业名称
同步链路
模式
最近运行
操作
销售订单 ODS 全量同步
bi_m8_sales_order → bi_ods.ods_sales_order
全量
未运行
全量执行
增量执行
运行记录
运行实例
批次号
作业
模式
状态
读写行数
尝试
创建时间
操作
暂无运行实例
共 0 条
10条/页
第4步重点操作 先手工运行一次
数脉平台 · 用户手册 V2.0

□□□ *36-04 □ 4*□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□



步骤图 36-05 第5步：核对无误后启用周期任务。红框标出核对无误后启用周期任务。
核对无误后启用周期任务，并在运行记录中查看批次和状态。

最终结果与效果
源表数据进入 ODS，页面显示批次号、行数、耗时和运行状态；重复执行不会重复新增订单。

操作提示： 同步只负责搬运，不负责业务汇总；字段清洗应在数据加工层完成。 “

当前演示环境说明
截图中的页面提示 “数据同步集成已关闭 ”，源库连接失败、 Doris ODS 连接失败、 ETL 调度不可用。该状态用于说明诊断信息如何呈现，部署环境启用后才会出现可执行的同步任务。

常见问题

执行按钮不可用。 作业可能已停用，或当前环境关闭了同步集成。

读取行数正常但写入为 0。 检查字段类型、目标表主键、 Doris 分区和写入权限。

重试后数据翻倍。 说明目标表没有正确使用批次幂等键；停止继续重试并联系数据开发核查。

开发人员参考
接口： GET /bi/etl/status、 GET/POST /bi/etl/jobs、 POST /bi/etl/jobs/{jobId}/runs、 POST /bi/etl/runs/{runId}/retry
权限： bi:etl:seatunnel
源表、目标表和字段映射由服务端白名单约束；执行过程使用内部令牌调用 SeaTunnel。
运行记录保存批次号、模式、读写行数、尝试次数和错误摘要，不在前端暴露连接口令。

数据加工

功能定位
数据加工负责把 ODS 原始数据清洗、标准化并逐层加工为 DWD 明细、 DWS 汇总和 ADS 应用表。数脉使用 Hop 2.19.0 承载固定加工管线，并把管线版本、运行记录、 Doris 产物预览和 DolphinScheduler 调度串在一起。

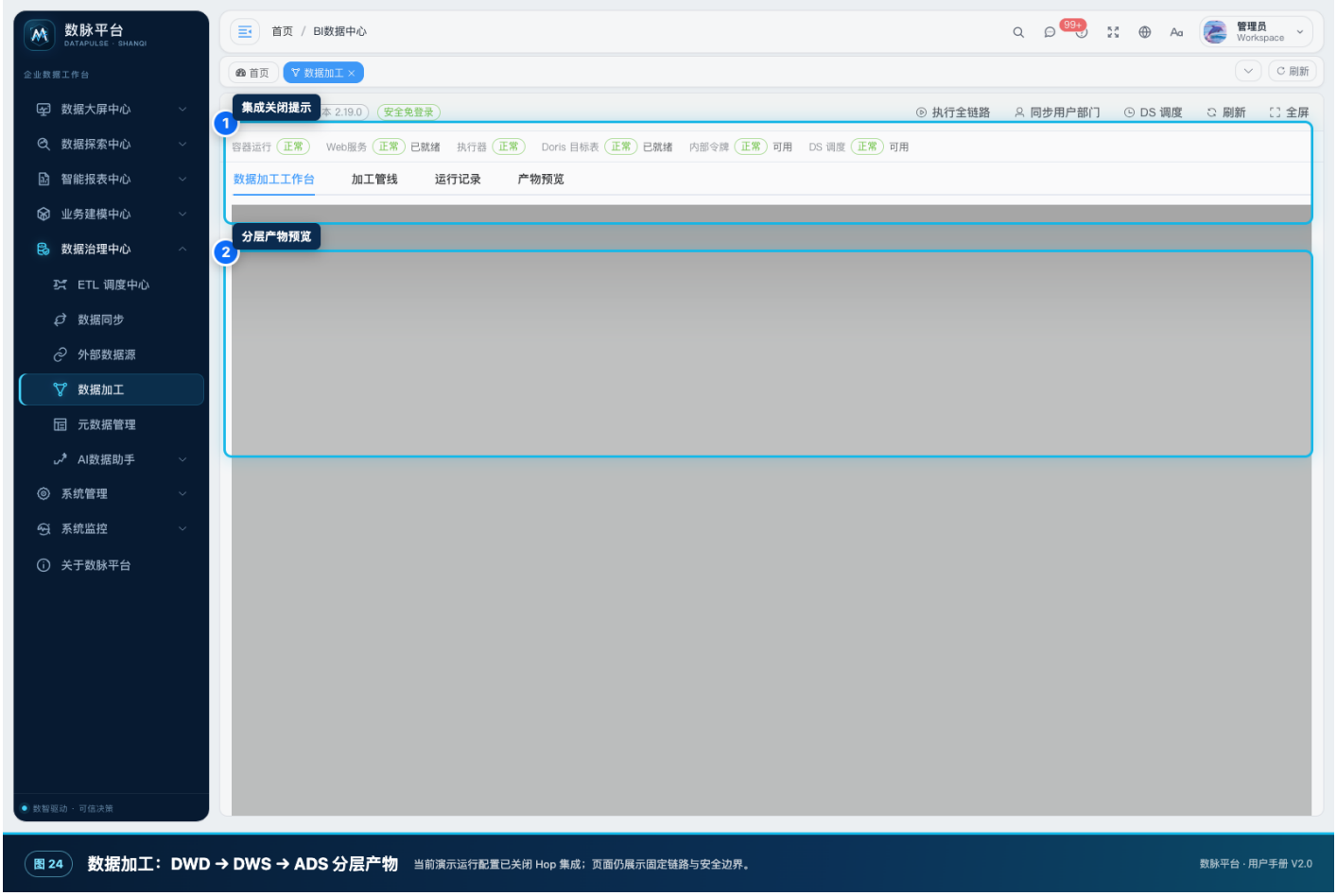


图 7-3 DWD → DWS → ADS 分层产物与集成状态

适用角色
数据开发人员、 ETL 管理员、数仓建模人员和数据运维人员。

页面入口

数据治理中心 → 数据加工（`/bi/etl/hop`）。

页面结构

页签/区域	说明
数据加工工作台	打开 Hop Web 工作台，查看或设计转换流程
加工管线	展示管线名称、分层、目标表、生效版本、最近运行和目标行数
运行记录	展示运行号、管线、版本、触发方式、状态、耗时和错误摘要
产物预览	按分层查看 Doris 目标表的样例行、字段和统计
顶部操作	执行全链路、同步用户部门、DS 调度、刷新、全屏

操作步骤

1. 先确认容器运行、Web 服务、执行器、Doris 目标表、内部令牌和 DS 调度状态。
2. 在“加工管线”中确认 ODS→DWD、DWD→DWS、DWS→ADS 的链路顺序和目标表。
3. 点击“执行全链路”前，确认上游 ODS 批次已完成，避免加工空数据。
4. 查看“运行记录”，逐个核对每个管线的读取行数、写入行数、耗时和失败原因。
5. 在“产物预览”检查主键唯一性、金额非负、日期范围、维度关联和汇总口径。
6. 需要每日自动运行时，通过“DS 调度”登记固定工作流；不要在页面上设置任意 Cron 表达式。
7. 调整管线后发布新版本并激活，保留旧版本以便回滚。

输入与输出

输入是 ODS/DWD/DWS 表、字段映射、清洗规则和管线版本；输出是 DWD 明细、DWS 汇总、ADS 应用表及运行记录。

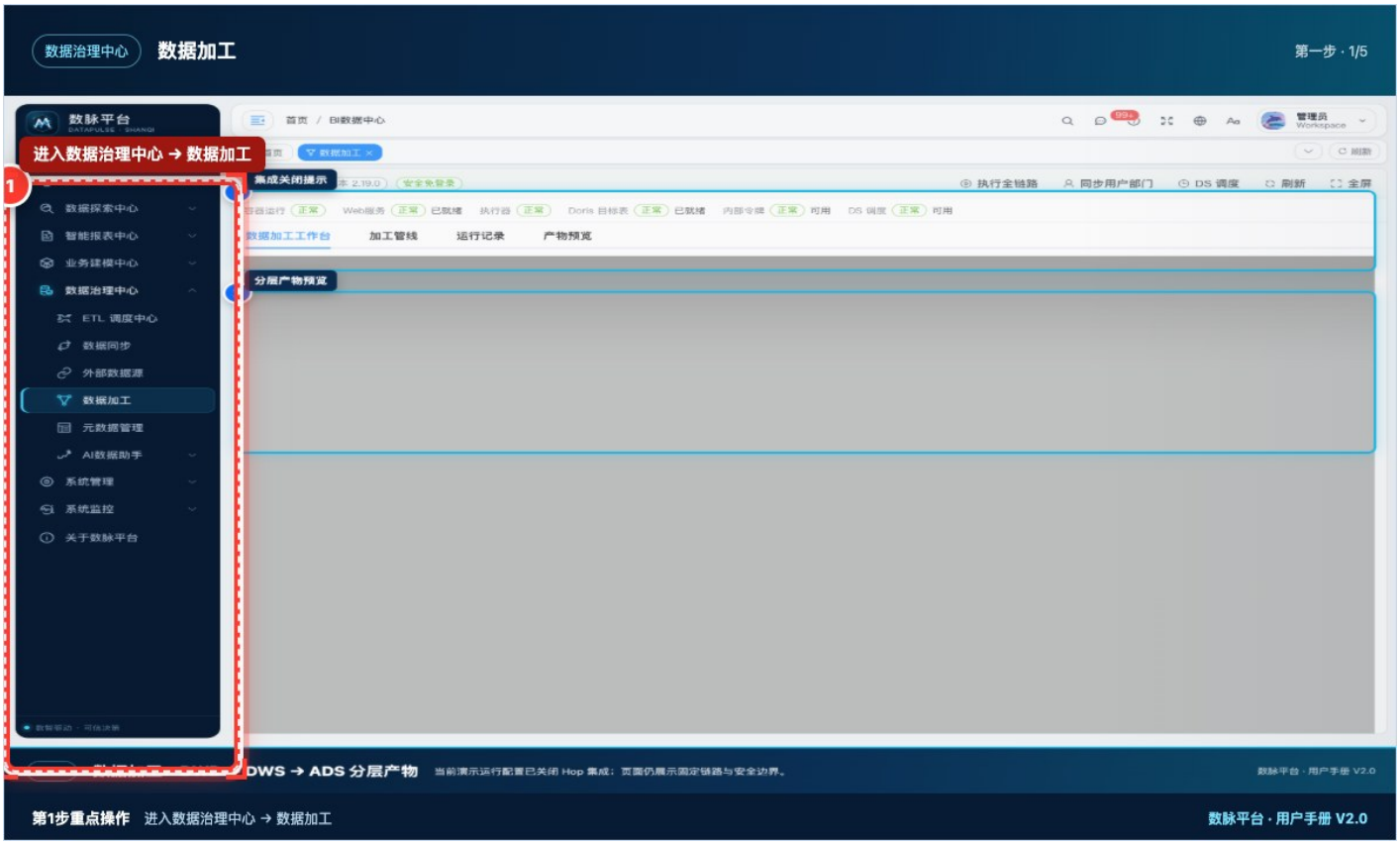
宏澜家电示例

订单清洗管线过滤取消订单、补齐客户与产品维度、统一金额单位并写入 DWD。按区域和品类汇总后形成 DWS 销售日汇总，再生成 ADS 部门销售应用表。若 DWD 行数比 ODS 少 2%

，先检查取消订单和异常过滤规则是否符合口径，而不是直接补行。

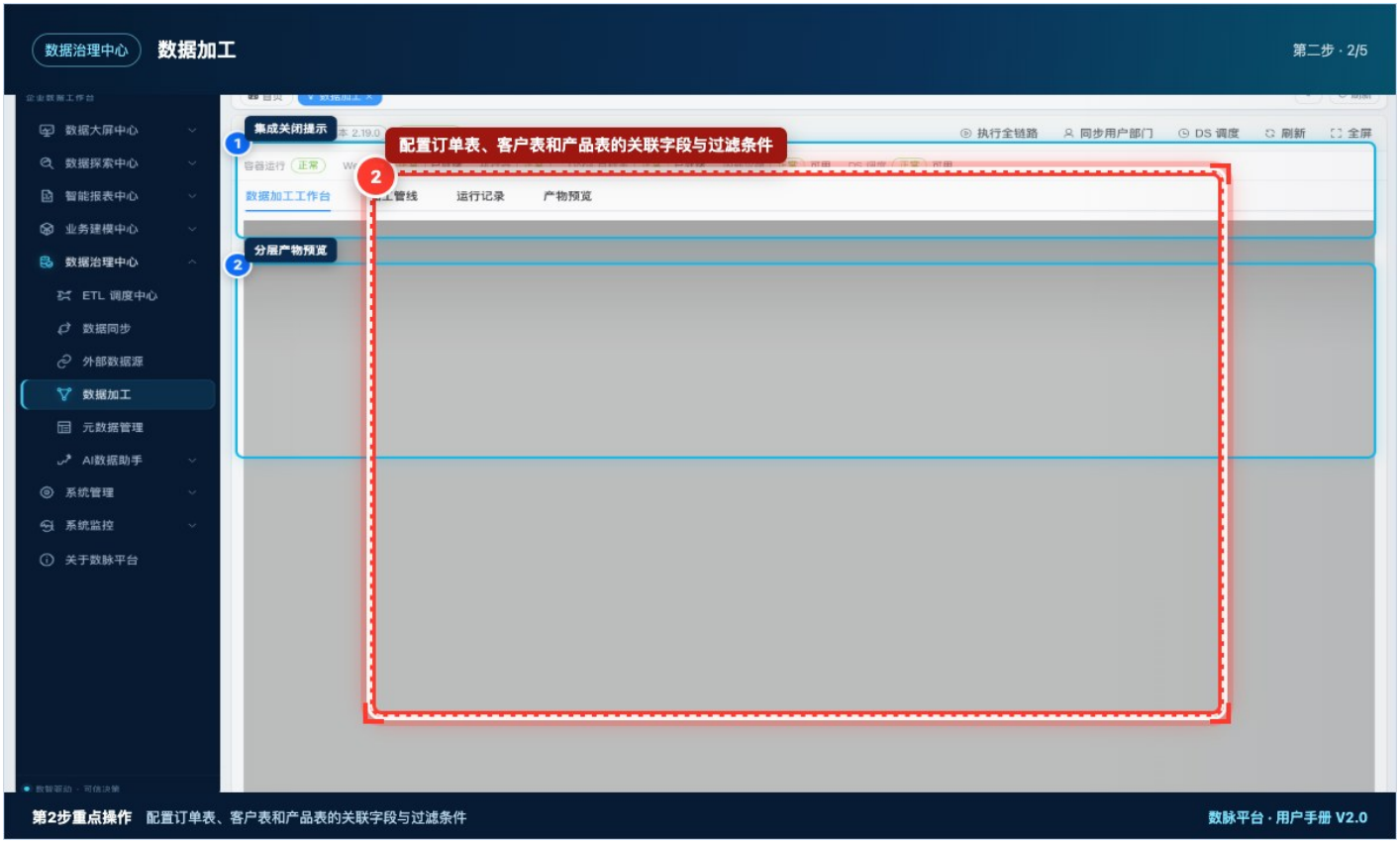
客户应用演示：把 ODS 订单加工为 DWD 明细
业务情境： ODS 中订单包含取消、测试和重复数据，需要清洗后进入 DWD。

第一步：进入 “数据治理中心 → 数据加工 ”



步骤图 37-01 第1步：进入 “数据治理中心 → 数据加工 ”。红框标出进入数据治理中心 → 数据加工。进入 “数据治理中心 → 数据加工 ”，选择 ODS → DWD 加工模板。

第二步：配置订单表、客户表和产品表的关联
字段与过滤条件



步骤图 37-02 第2
步：配置订单表、客户表和产品表的关联字段与过滤条件。红框标出配置订单表、客户表和产品表的关联字段与过滤条件。

配置订单表、客户表和产品表的关联字段与过滤条件。

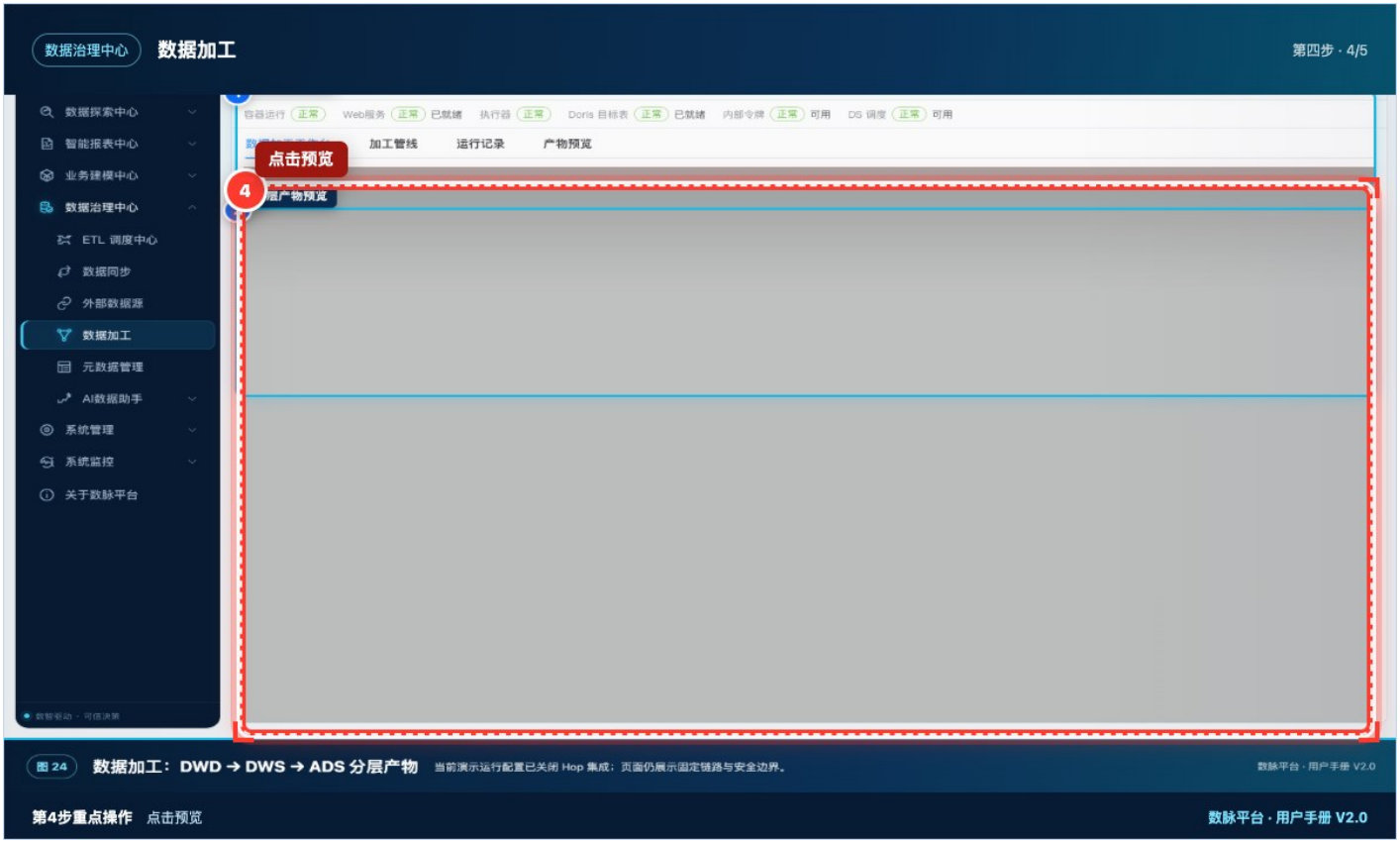
第三步：处理取消订单、重复订单、空客户、金额格式和日期标准化



步骤图 37-03 第3
步：处理取消订单、重复订单、空客户、金额格式和日期标准化。红框标出处理取消订单、重复订单、空客户、金额格式和日期标准化。

处理取消订单、重复订单、空客户、金额格式和日期标准化。

第四步：点击预览



步骤图 37-04 第4步：点击预览。红框标出点击预览。

点击预览，检查输入行数、输出行数和关键字段样例。

第五步：保存并运行



步骤图 37-05 第5步：保存并运行。红框标出保存并运行。
保存并运行，确认 DWD 产物可被 DWS/ADS 继续加工。

最终结果与效果
形成干净、结构化、可追溯的 DWD 订单明细，运行记录保留输入输出行数和批次。

操作提示： 加工前先确认业务规则，尤其是取消、退货和折扣的含义。 “

当前演示环境说明
截图显示 Hop 集成处于 DISABLED，容器、 Web 服务、执行器和 Doris 目标表状态异常。页面仍展示固定的分层链路和安全边界，用于说明启用后的操作路径。

常见问题

全链路执行后中间层有数、应用层为空。 检查 DWS/ADS 的日期分区、汇总维度和调度依赖。

Doris 目标表找不到。 先完成模型发布或建表，再执行加工；不要在 Hop 中临时创建任意表。

运行成功但金额异常。 检查币种、含税 /未税、折扣和退货口径，运行成功不代表业务口径正确。

开发人员参考
接口： /bi/etl/hop/status、 /launch、 /pipelines、 /pipelines/{id}/versions、 /runs/chain
、 /schedule、 /preview
权限： bi:etl:hop
管线目标表与文件路径由服务端白名单固定；版本发布、激活、运行和回滚均有独立记录。
Hop 账号凭据不进入浏览器，使用一次性票据交换受限会话。

外部数据源

功能定位
外部数据源用于接入不适合直接连库的第三方系统，例如电商平台、企业微信、开放 API、SaaS 订单接口。系统通过受控 HTTP 出口读取分页或游标数据，写入指定 Doris
目标表，并记录批次、重试、审计和脱敏原因。

图 7-4 外部源配置、同步作业、运行实例与安全审计

适用角色
ETL 管理员、集成开发人员、数据运维人员和审计人员。

页面入口

数据治理中心 → 外部数据源（`/bi/etl/external`）。

页面结构

页签	作用
外部源	源名称、类型、接口地址、目标表、鉴权方式、状态和编辑入口
同步作业	作业名称、外部源、全量/增量模式、自动重试和最近运行
运行实例	批次号、作业、状态、读写行数、DS 实例和重试
安全审计	事件、结果、操作人、脱敏原因和来源IP

操作步骤

1. 点击“新建外部源”，选择模板：通用 REST 分页、企业微信游标或开放 API。
2. 填写源名称、接口地址、请求方法、目标表和分页参数；接口地址只允许服务端出口白名单中的域名。
3. 选择鉴权方式（如 HEADER 或 QUERY），凭据使用 `env:` 或文件引用，页面只显示“已配置”，不回显明文。
4. 保存后测试连通性，确认返回结构、分页字段、游标路径和错误码。
5. 新建同步作业，设置全量或增量、最大重试次数和启停状态。
6. 执行后查看运行实例的读取行数、写入行数和失败原因；重试保持批次号。
7. 在安全审计中核对每次配置、执行和脱敏事件。

输入与输出

输入是接口地址、鉴权引用、分页规则、字段映射和目标表；输出是写入 Doris 的外部数据、批次记录、审计事件和脱敏摘要。

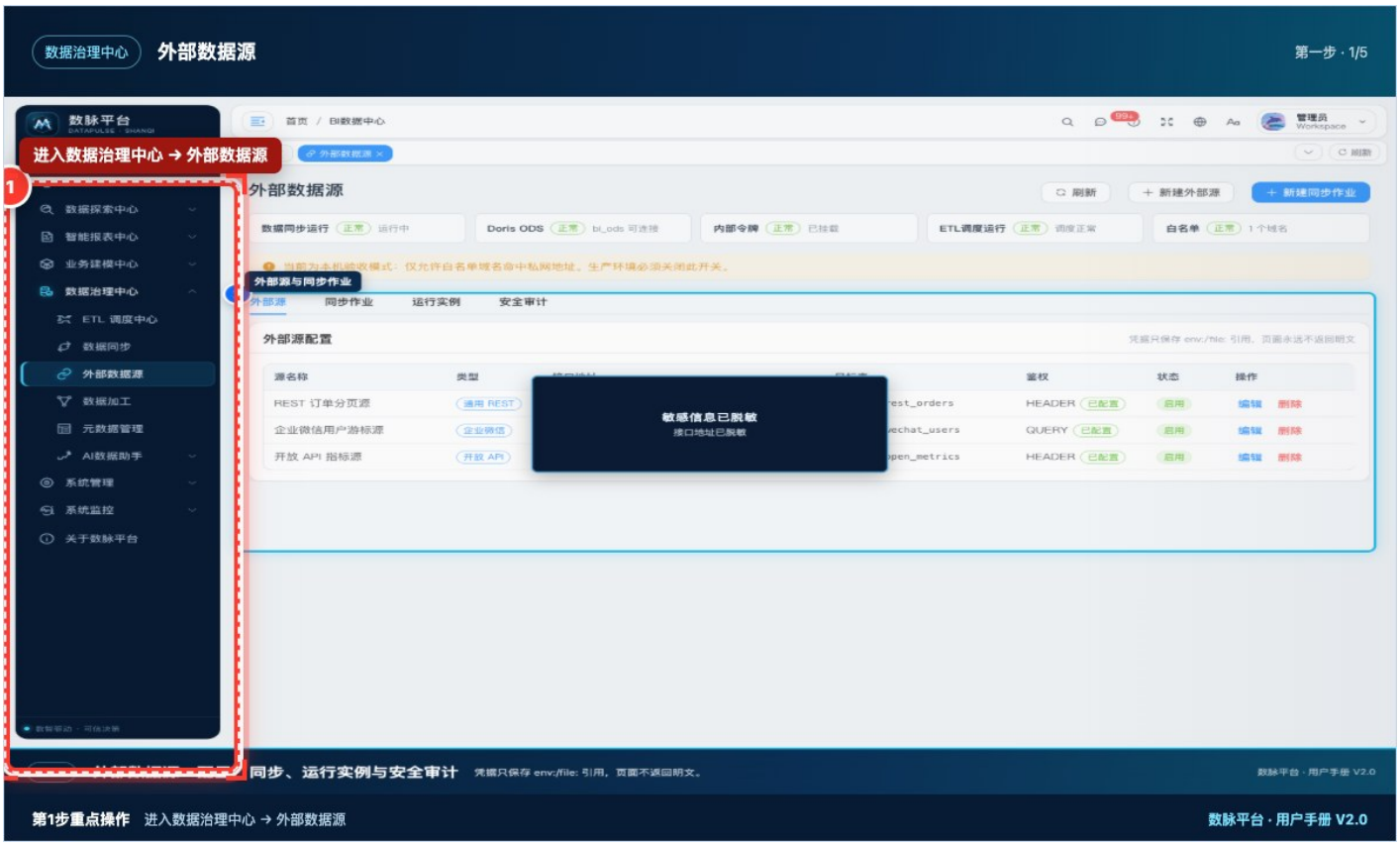
宏澜家电示例

接入某电商平台的订单接口：使用 REST 分页，按更新时间增量拉取，订单号作为幂等键写入
ods_ecommerce_order。凭据由环境变量注入，页面只显示 “HEADER 已配置 ”
。若平台限流，作业按指数退避重试并保留原批次号。

客户应用演示：接入宏澜 ERP

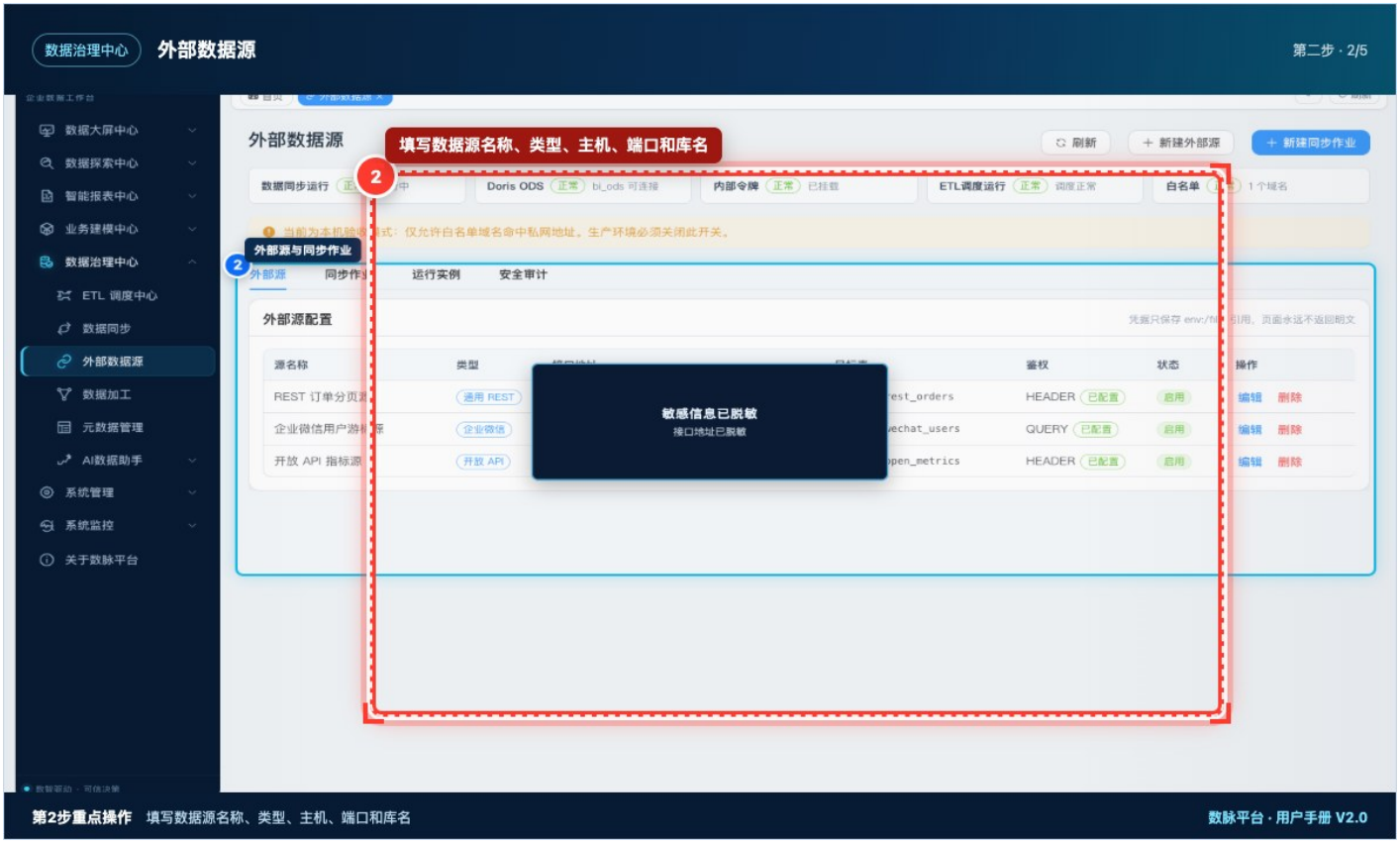
只读数据库
业务情境： 数据管理员需要把外部 ERP 作为后续同步和建模的数据源。

第一步：进入 “数据治理中心 → 外部数据源 ”



步骤图 38-01 第1步：进入 “数据治理中心 → 外部数据源 ”。红框标出进入数据治理中心 → 外部数据源。
进入 “数据治理中心 → 外部数据源 ”，点击新增。

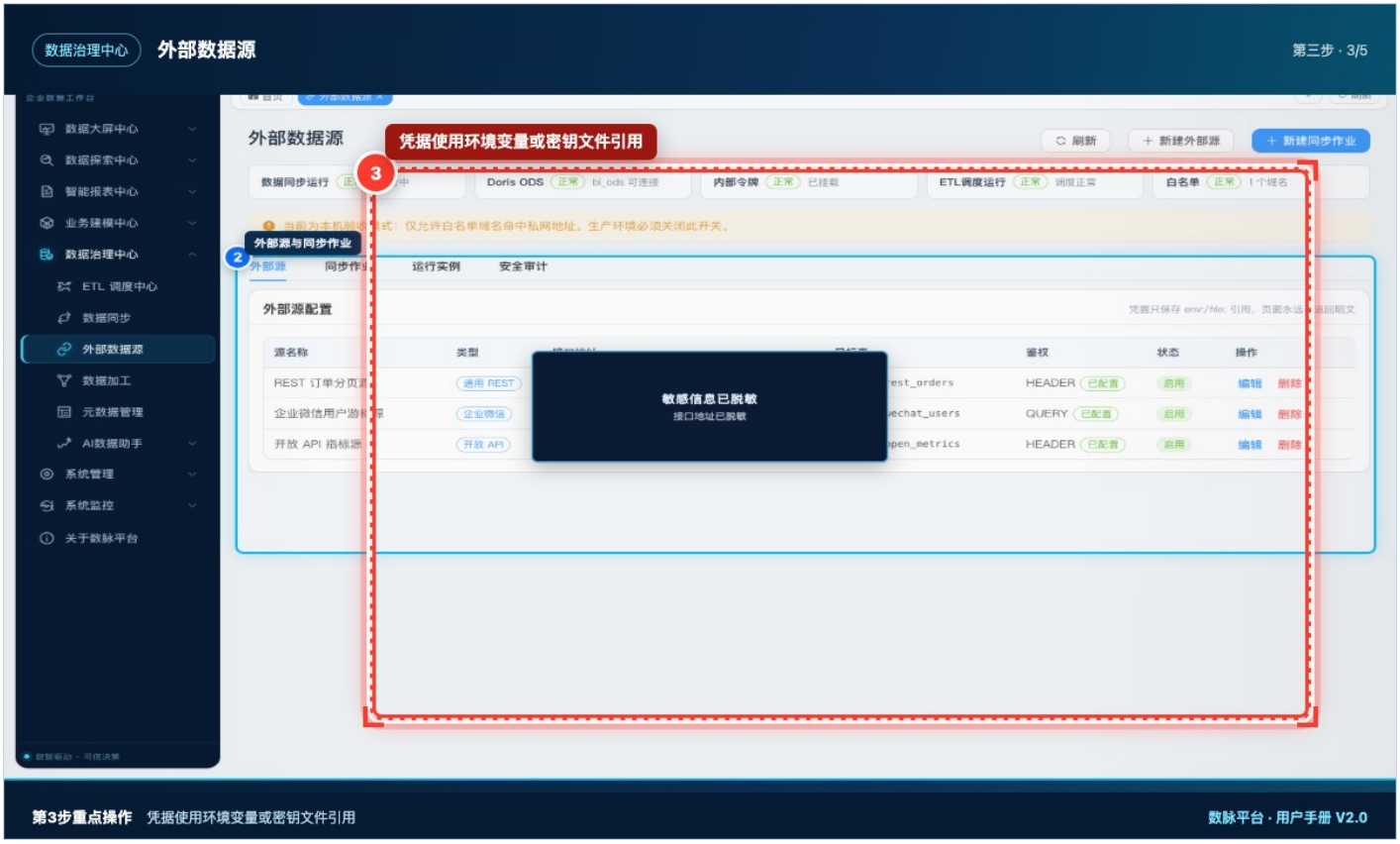
第二步：填写数据源名称、类型、主机、端口
和库名



步骤图 38-02 第2
步：填写数据源名称、类型、主机、端口和库名。红框标出填写数据源名称、类型、主机、端口和库名。

填写数据源名称、类型、主机、端口和库名。

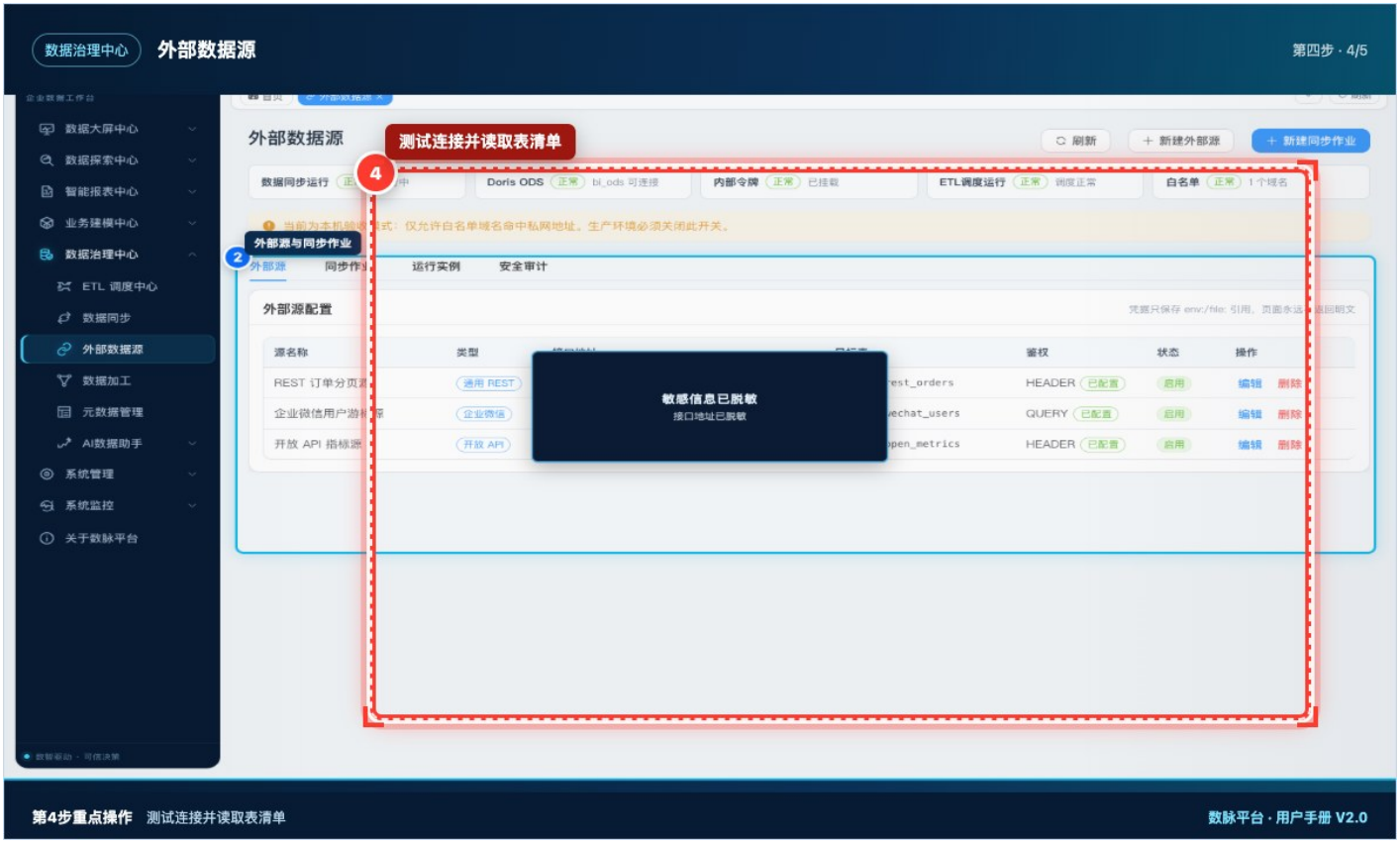
第三步：凭据使用环境变量或密钥文件引用



步骤图 38-03 第3步：凭据使用环境变量或密钥文件引用。红框标出凭据使用环境变量或密钥文件引用。

凭据使用环境变量或密钥文件引用，不填写明文到页面说明中。

第四步：测试连接并读取表清单



步骤图 38-04 第4步：测试连接并读取表清单。红框标出测试连接并读取表清单。

测试连接并读取表清单，确认只读账号权限边界。

第五步：保存后执行一次连接诊断



步骤图 38-05 第5步：保存后执行一次连接诊断。红框标出保存后执行一次连接诊断。

保存后执行一次连接诊断，并把数据源负责人和用途登记完整。

最终结果与效果
外部数据源登记成功，后续同步作业可按白名单引用；页面不回显密码，操作进入审计日志。

操作提示： “
连接失败先检查网络白名单、账号权限和密钥引用；不要把真实连接串写入文档。

当前演示环境说明
截图显示 “外部数据源集成已关闭 ”，接口地址已脱敏，白名单为 0 个域名。该页面展示配置结构和审计方式；实际调用前必须由管理员开启集成并配置允许访问的域名。

安全边界
禁止接入任意公网地址，禁止通过页面探测内网服务。
凭据只保存引用，不返回明文；日志和审计不得记录 Token、Cookie 或完整响应。
目标表必须在白名单内，字段映射必须显式声明。
删除数据源前先停用作业并确认历史批次和下游依赖。

常见问题
保存时提示域名不在白名单。 由管理员在服务端出口策略中登记域名，不要在页面反复换写法绕过。

运行成功但字段为空。 检查 JSON 路径、分页游标和字段映射，先用小页数验证。

凭据显示 “未配置 ”。 环境变量或文件引用没有加载，需管理员检查部署配置，不要把密钥直接填到浏览器。

开发人员参考
接口： /bi/ext/sources、 /jobs、 /runs、 /audits、 /options、 /status
权限： bi:etl:ext
使用受控出口网关，服务端校验目标域名、请求方法、响应大小和字段白名单。
审计只保存事件、结果和脱敏原因；凭据引用不写入数据库明文。

元数据管理

功能定位
元数据管理把 Doris 表、字段、中文注释、 dbt
模型、质量结果和上下游血缘集中到一个可检索目录。它回答三个问题：平台上有什么数据、字段是什么意思、这个指标从哪里来又流向哪里。

图 7-5 治理通道诊断、 Doris 元数据目录与字段血缘入口

适用角色
数据架构师、数据管理员、指标管理员、开发人员和审计人员。

页面入口

数据治理中心 → 元数据管理（`/bi/etl/metadata`）。

页面结构

区域	说明
治理通道诊断	元数据服务、OM 授权、dbt 环境、dbt 产物、Doris 采集、质量同步、ETL 调度服务
元数据目录	仅展示服务端白名单 Schema，可搜索表名或注释
字段与注释	查看字段类型、中文名、描述、敏感级别和是否可用于分析
数据血缘	查看表级和字段级上下游关系
dbt 模型	查看模型、依赖、最近运行和质量结果
质量与运行	查看质量测试、运行状态和失败原因
调度实例	查看与元数据刷新相关的工作流实例

操作步骤

1. 先查看治理通道诊断。红色状态表示该能力不可用，需管理员处理。
2. 在元数据目录选择 Schema，输入表名或中文注释查询。
3. 点击“字段”查看字段清单、类型和注释，确认口径后再用于建模或分析。
4. 点击“血缘”查看上游来源、下游模型和看板引用，评估变更影响。
5. 需要更新目录时，由管理员执行“元数据采集”，选择 Doris、dbt、质量结果或全量刷新。
6. 需要验证 dbt 时，使用“dbt 命令”执行 debug、run、test 或 docs generate。
7. 需要周期性刷新时，通过“执行受控调度”登记固定工作流，并核对调度实例结果。

输入与输出

输入是 Doris Schema、dbt manifest、质量结果和调度状态；输出是可搜索目录、字段注释、血缘图、dbt 模型清单和采集任务记录。

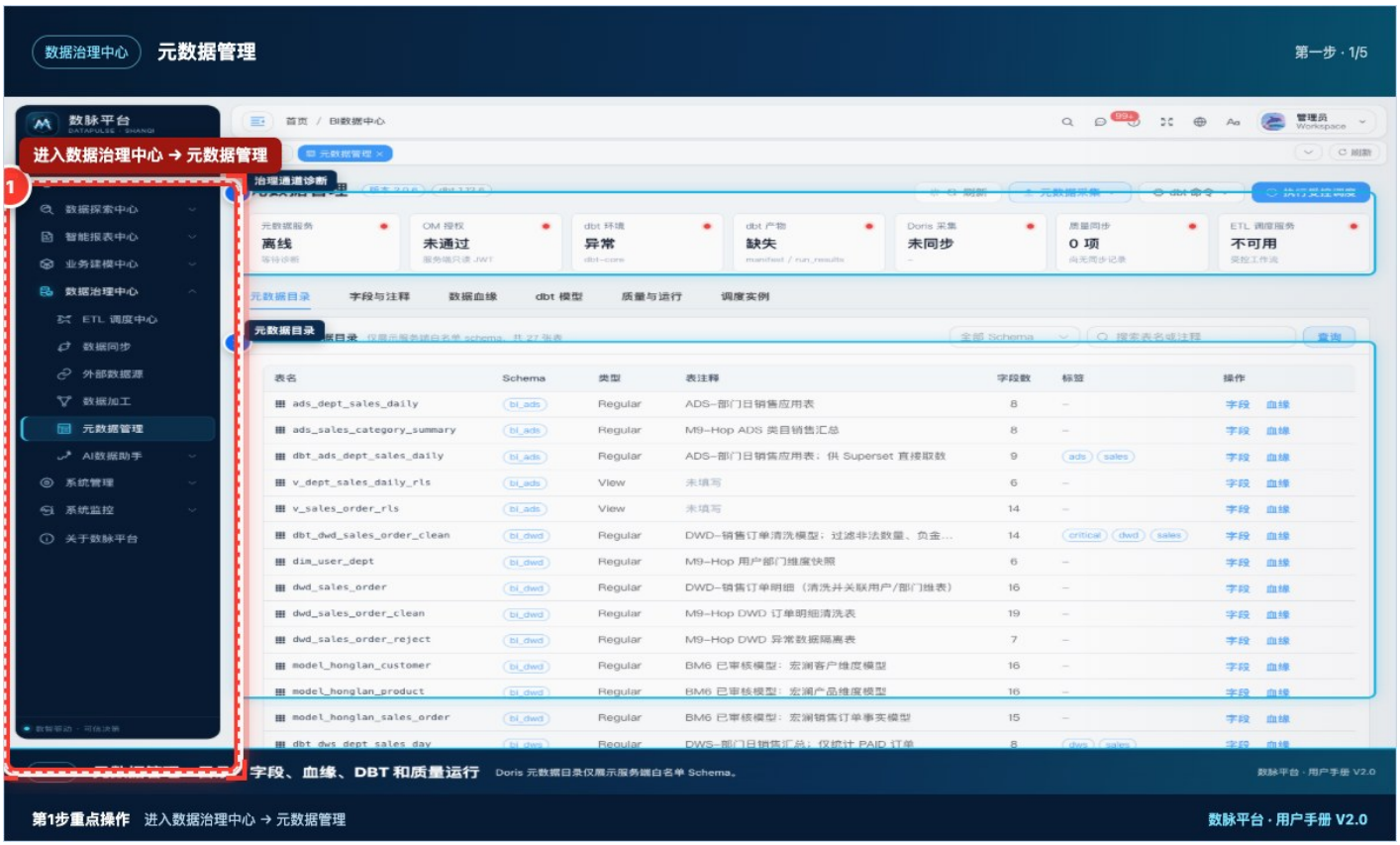
宏澜家电示例

分析师发现 “区域销售额 ”与财务口径不一致，在元数据管理中打开 ads_dept_sales_daily 的血缘：上游是 dws_sales_category_summary 和 dwd_sales_order_clean，再向上到 ERP 订单表。通过字段注释确认 “ 净销售额 ”已扣除退货，从而定位差异来自财务使用含税收入口径。

客户应用演示：追踪 “销售额 ”

字段的来源与影响
业务情境： 业务人员质疑销售额口径，需要确认字段来自哪里、被哪些资产使用。

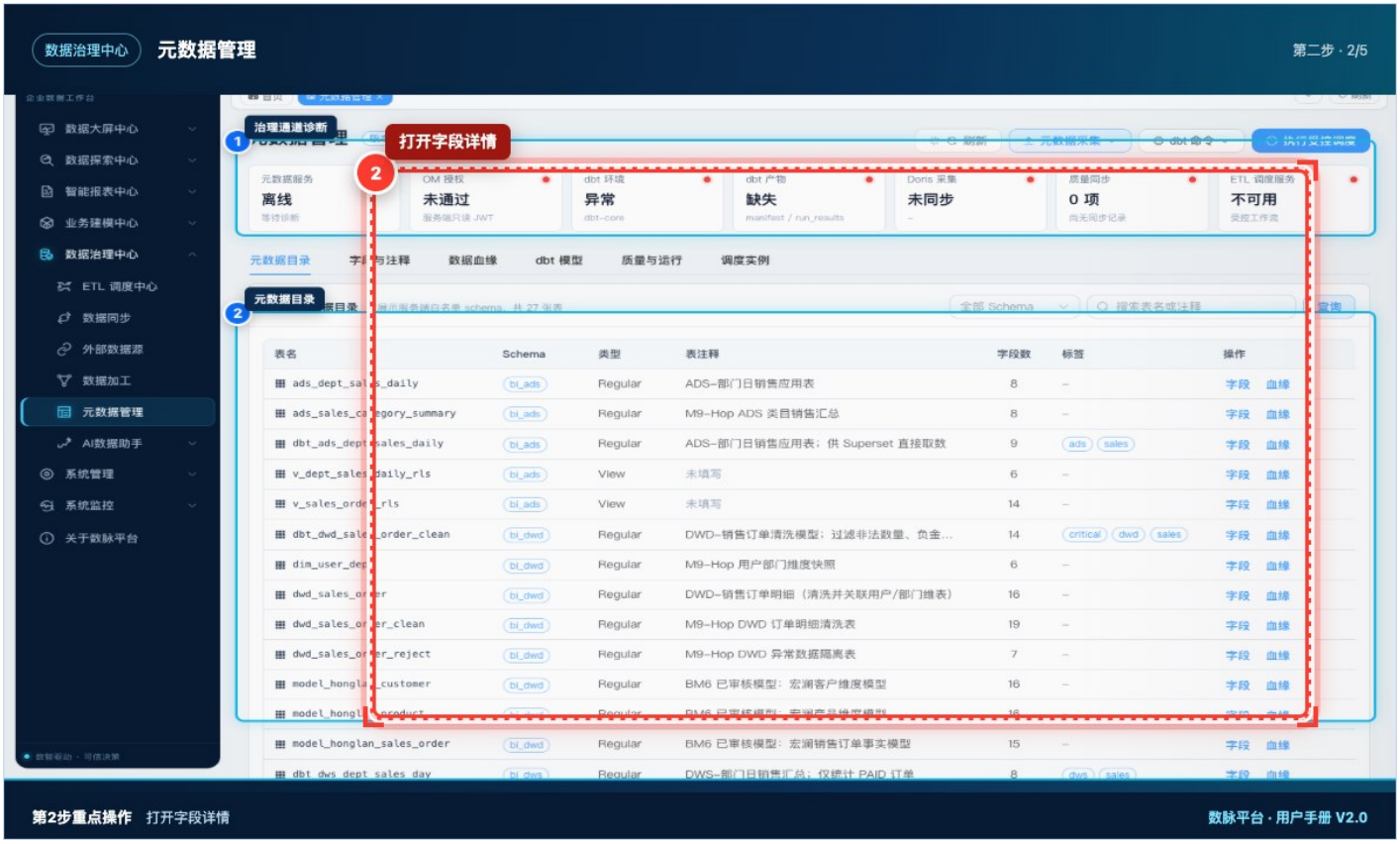
第一步：进入 “数据治理中心 → 元数据管理 ”



步骤图 39-01 第1步：进入 “数据治理中心 → 元数据管理 ”。红框标出进入数据治理中心 → 元数据管理。

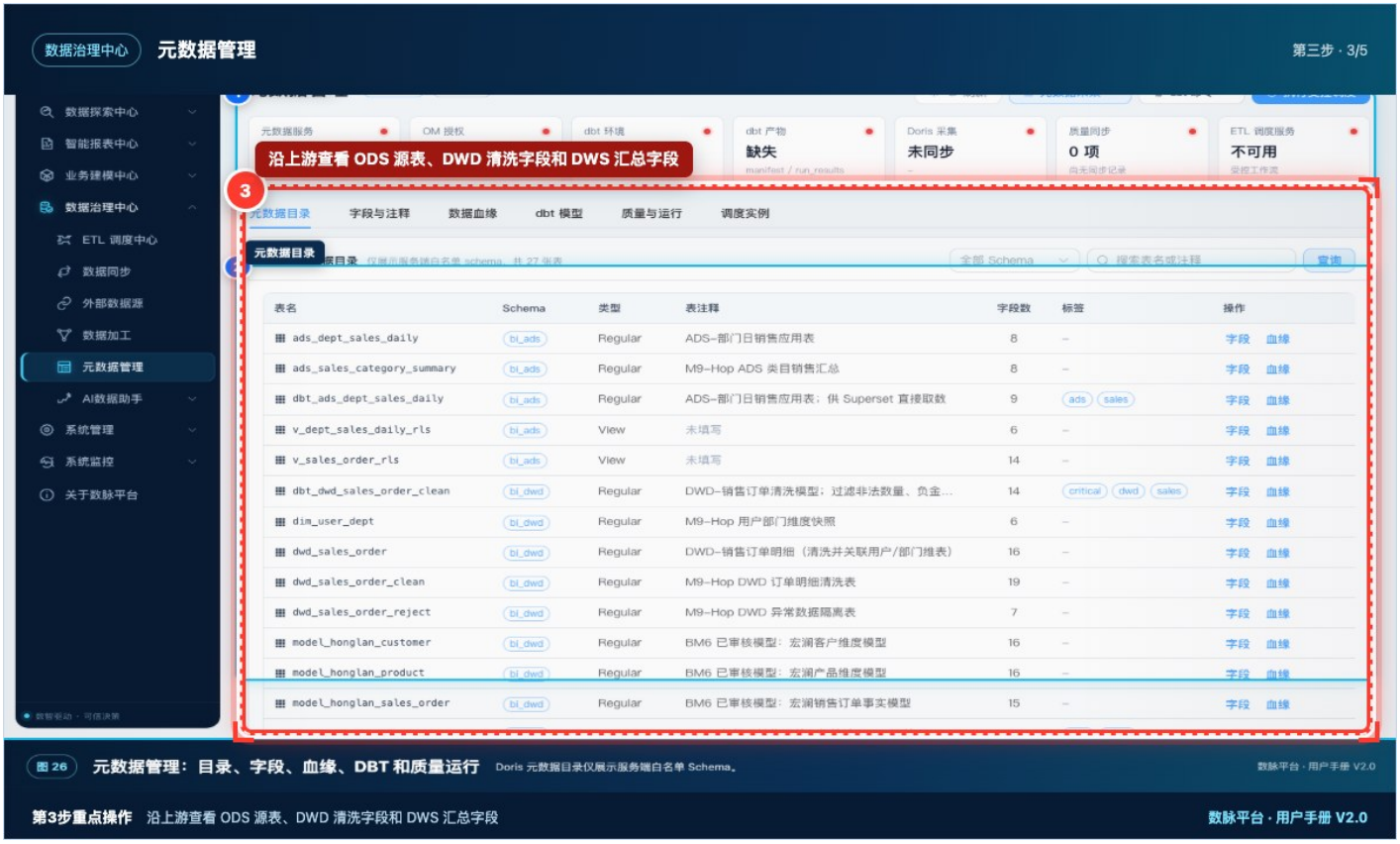
进入 “数据治理中心 → 元数据管理 ”，搜索销售额字段。

第二步：打开字段详情



步骤图 39-02 第2步：打开字段详情。红框标出打开字段详情。
打开字段详情，查看所属表、数据类型、注释和负责人。

第三步：沿上游查看 ODS 源表、 DWD 清洗字段和 DWS 汇总字段

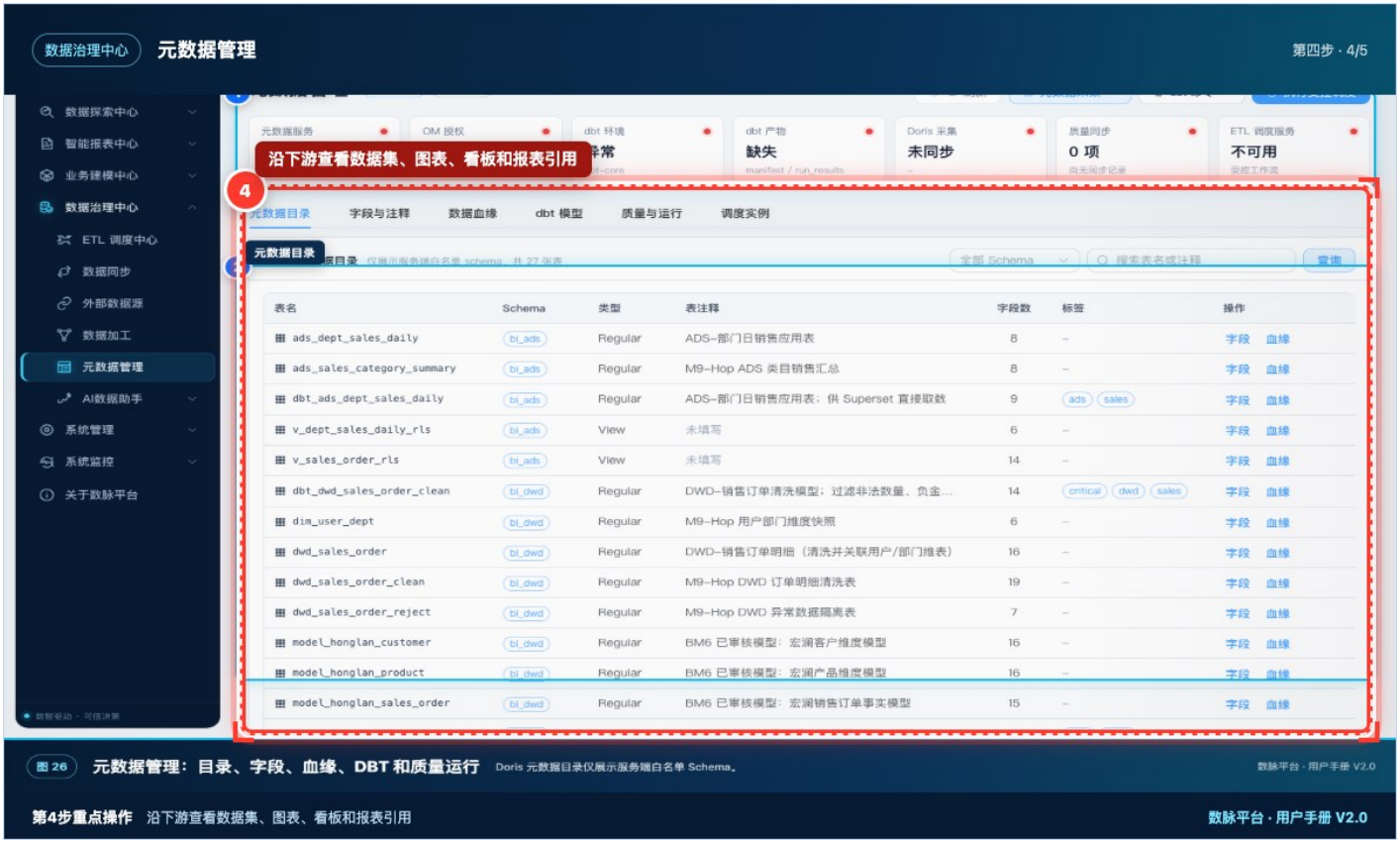


□□□ *39-03 □ 3□□□□□□□ ODS □□□ DWD □□□□□ DWS □□□□□□□□□□□□□□□ ODS □□□ DWD □□□□□ DWS □□□□□*

□□□□□ ODS □□□ DWD □□□□□ DWS □□□□□

□□□□□□□□□□□□□□□□□□□□□

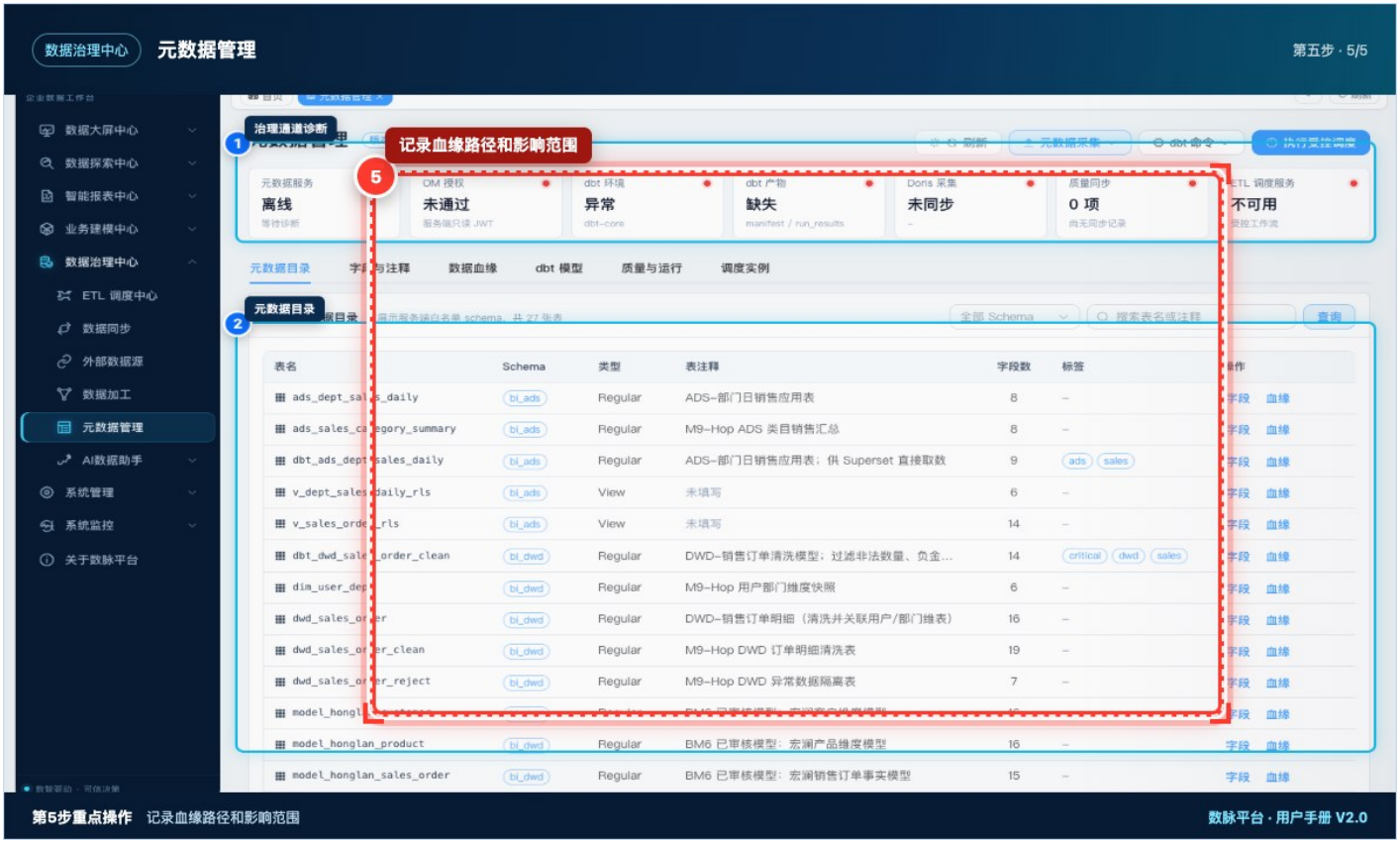
□□□



步骤图 39-04 第4
步：沿下游查看数据集、图表、看板和报表引用。红框标出沿下游查看数据集、图表、看板和报表引用。

沿下游查看数据集、图表、看板和报表引用。

第五步：记录血缘路径和影响范围



步骤图 39-05 第5步：记录血缘路径和影响范围。红框标出记录血缘路径和影响范围。
记录血缘路径和影响范围，如需修改先通知下游负责人。

最终结果与效果
页面显示字段的上游来源、处理过程和下游资产，能够判断一次口径变更会影响哪些报表。

操作提示： 血缘用于定位影响，不替代数据质量校验；修改前仍要抽样对账。 “

当前演示环境说明
截图显示元数据服务离线、 OM 授权未通过、 dbt 环境异常、 dbt 产物缺失、 Doris 采集已同步、质量同步为 0 项、 ETL 调度可用。页面同时展示 27 张白名单表。该状态用于演示诊断，不能据此认定生产血缘完整。

常见问题

搜不到表。 可能不在白名单 Schema，或尚未执行元数据采集。

血缘图只有表、没有字段。 需要 dbt manifest 和字段级依赖解析；先检查 dbt 产物是否生成。

表注释为空。 在模型或建表阶段补充 COMMENT，再重新采集，不要只在前端手工改显示。

开发人员参考
接口： /bi/metadata/status、 /catalog、 /table、 /lineage、 /dbt/models、 /dbt/runs、 /ingest、 /schedule
权限： bi:metadata:view，采集、 dbt 和调度为独立细粒度权限。
仅允许白名单 Schema；采集任务异步执行并记录输入摘要、结果和错误码。
血缘来自 dbt manifest、 Doris 元数据和登记关系的组合，不靠前端猜测。

AI 同步生成器

功能定位
AI 同步生成器根据白名单源表、字段注释和业务描述，生成 SeaTunnel 同步候选配置。它把 “ 从哪张表同步到哪张 ODS 表、保留哪些字段、按什么方式增量 ”写成结构化候选，但不会自动执行。

图 7-6 自然语言输入、候选草稿与源字段说明

适用角色
ETL 管理员、集成开发人员和数据开发人员。

页面入口

数据治理中心 → AI 数据助手 → AI 同步生成器（`/bi/etl/ai/sync`）。

前置条件

- AI 服务正常，当前演示环境使用 `DeepSeek-V4.1-Flash`。
- 目标源表在服务端白名单内，并已完成元数据采集。
- DolphinScheduler 登记通道可用，当前账号有 AI 流水线权限。

操作步骤

1. 选择服务端模板，例如“销售订单全量同步”。
2. 选择白名单源表，页面会加载字段名、类型和中文注释。
3. 填写方案标题和业务描述。描述应说明同步频率、目标 Schema、需要保留的字段和过滤条件。
4. 点击“AI 生成候选流水线”，等待返回草稿编号、模型名和候选内容。
5. 在“结构化预览”检查源表、目标表、字段映射、主键和增量策略；必要时切换“JSON 编辑”修正白名单内字段。
6. 点击“确定性校验”，处理阻断错误和告警。
7. 提交人工审核，审核通过后绑定内容摘要；最后“登记发布”到 DolphinScheduler 门禁。
8. 发布后仍需按正常同步流程执行和核对数据。

输入与输出

项目	说明
输入	模板、白名单源表、方案标题、最多 1200 字业务描述
输出	同步候选草稿、字段映射、校验结果、审核记录、版本和发布摘要
状态	待校验、已校验、已审核、已发布、已驳回

宏澜家电示例

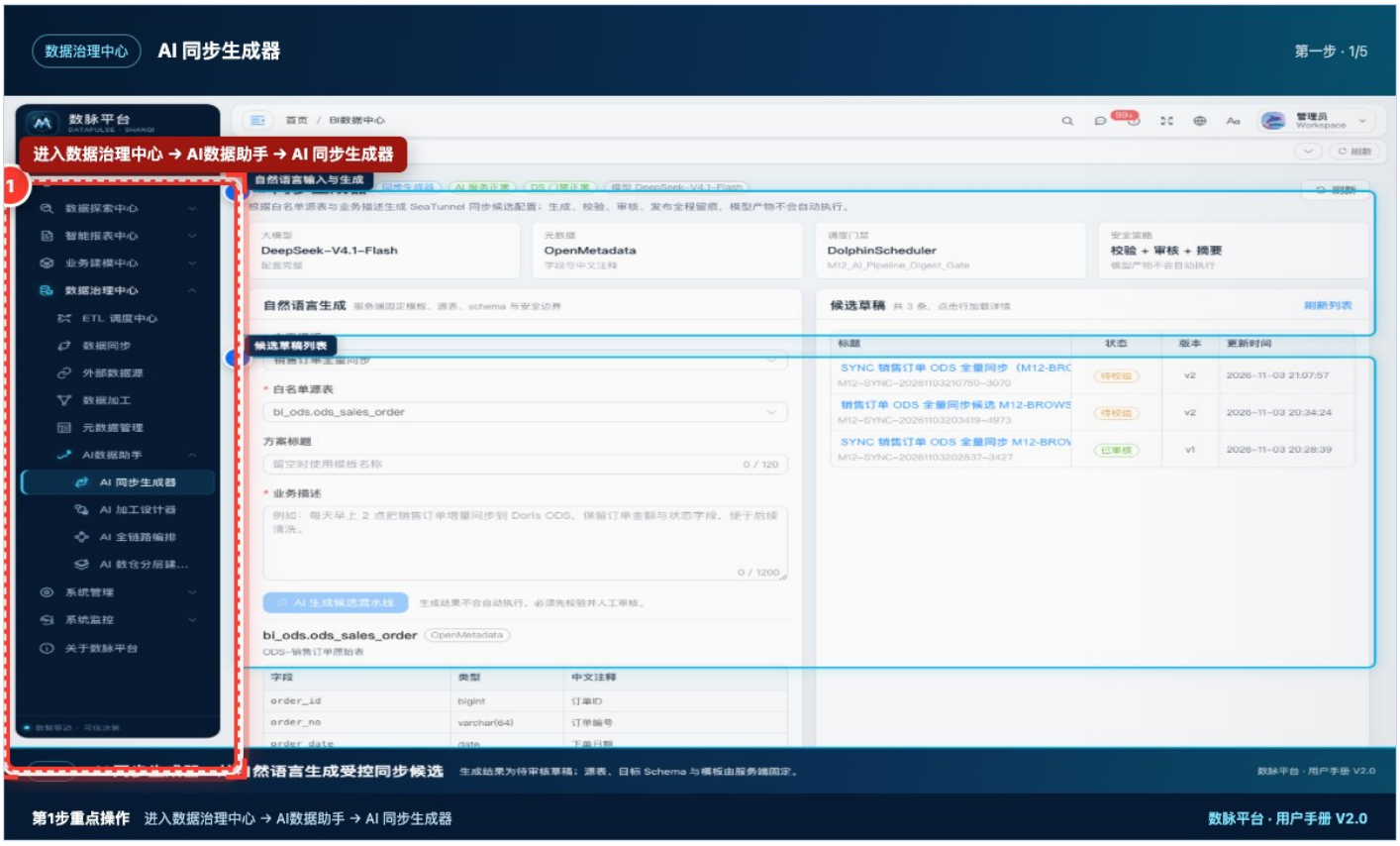
业务描述：“每天 02:00 把销售订单增量同步到 Doris ODS，保留订单金额、状态、客户、产品和更新时间，已取消订单也保留以便审计。”AI 生成订单源表到

`ods_sales_order`

AI

AI

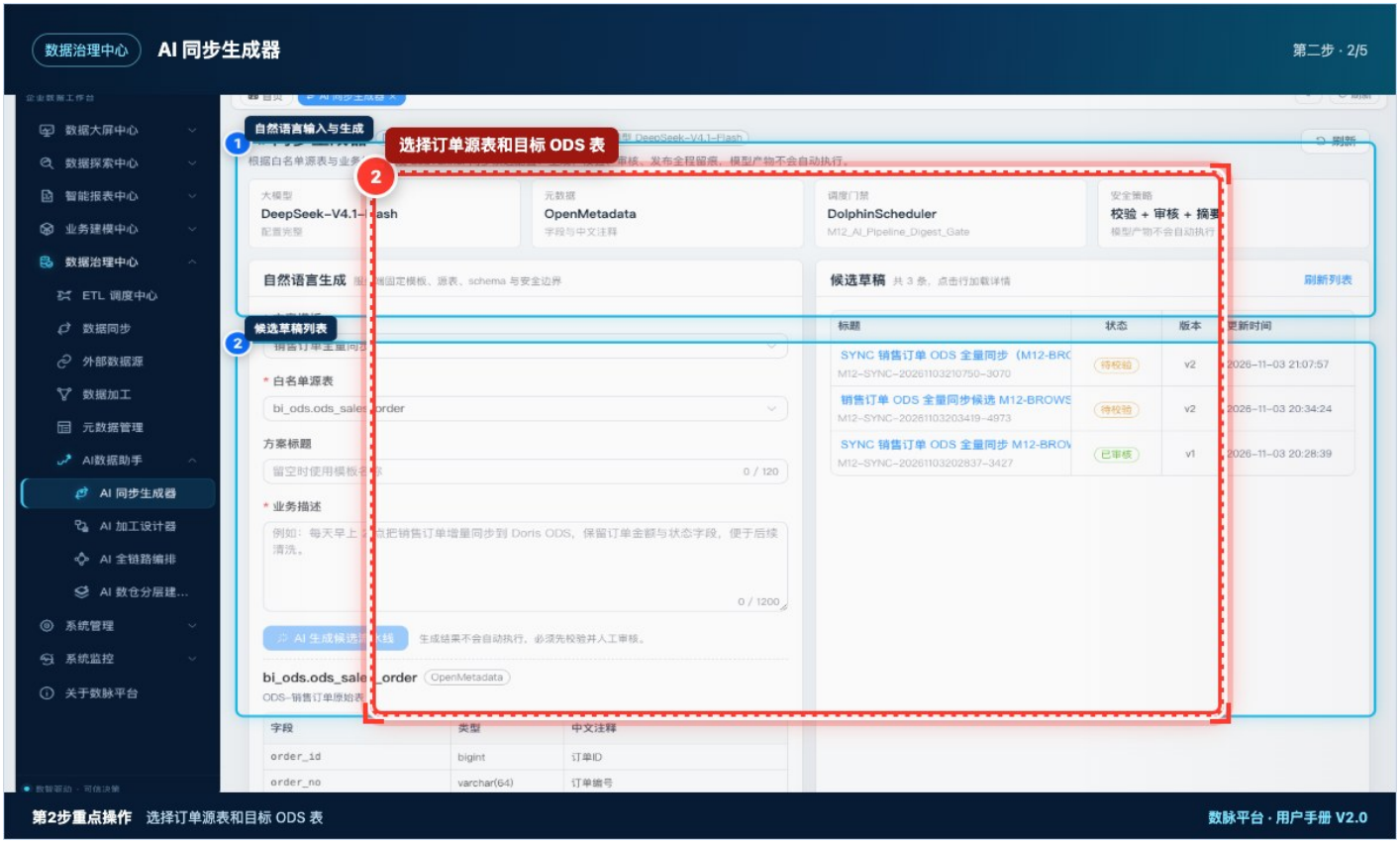
“ → AI → AI ”



40-01 1 “ → AI → AI ”→ AI → AI

“ → AI → AI ”

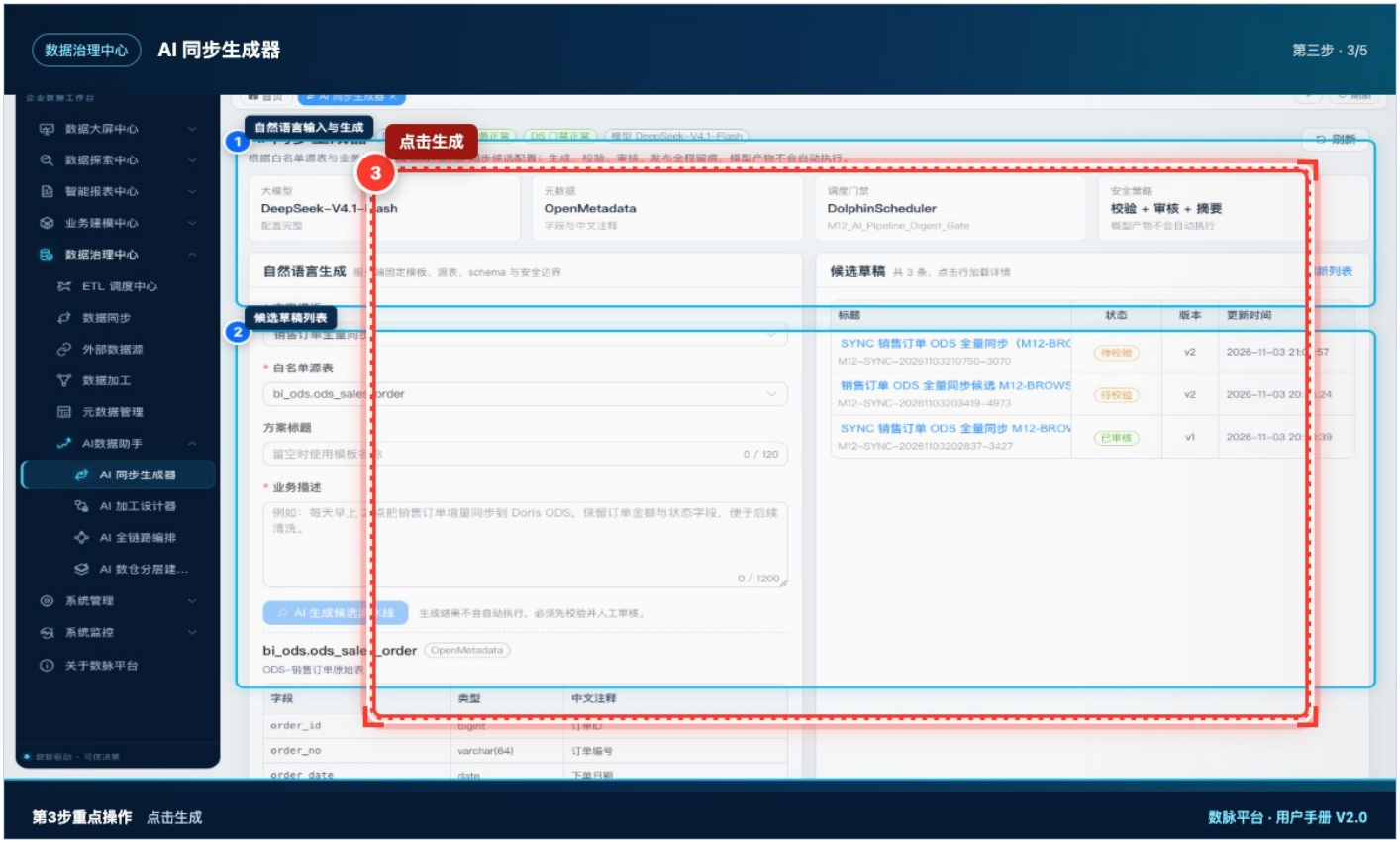
ODS



步骤图 40-02 第2步：选择订单源表和目标 ODS 表。红框标出选择订单源表和目标 ODS 表。

选择订单源表和目标 ODS 表，填写同步频率、增量字段和业务说明。

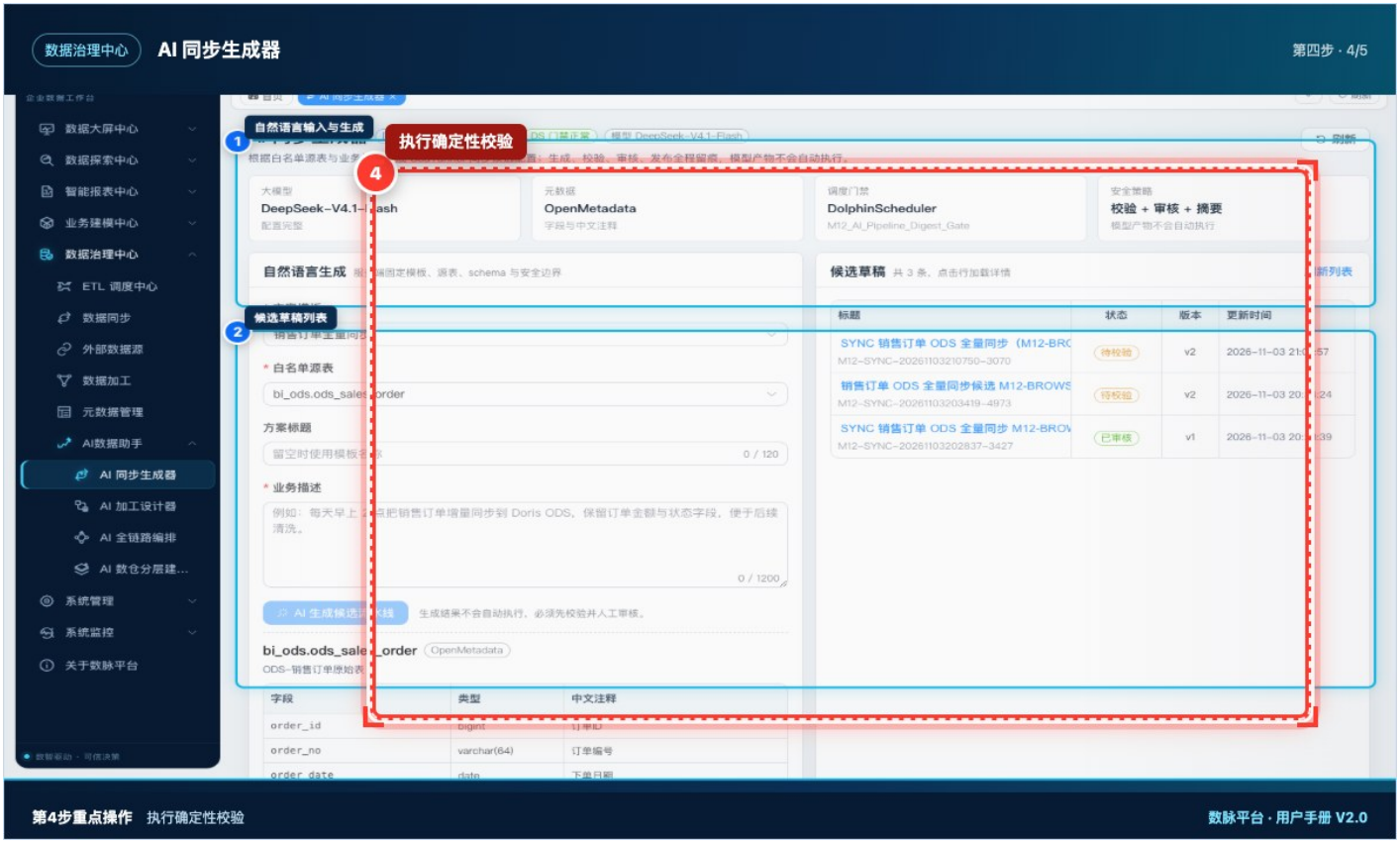
第三步：点击生成



步骤图 40-03 第3步：点击生成。红框标出点击生成。

点击生成，查看源字段、目标字段、过滤条件和安全边界。

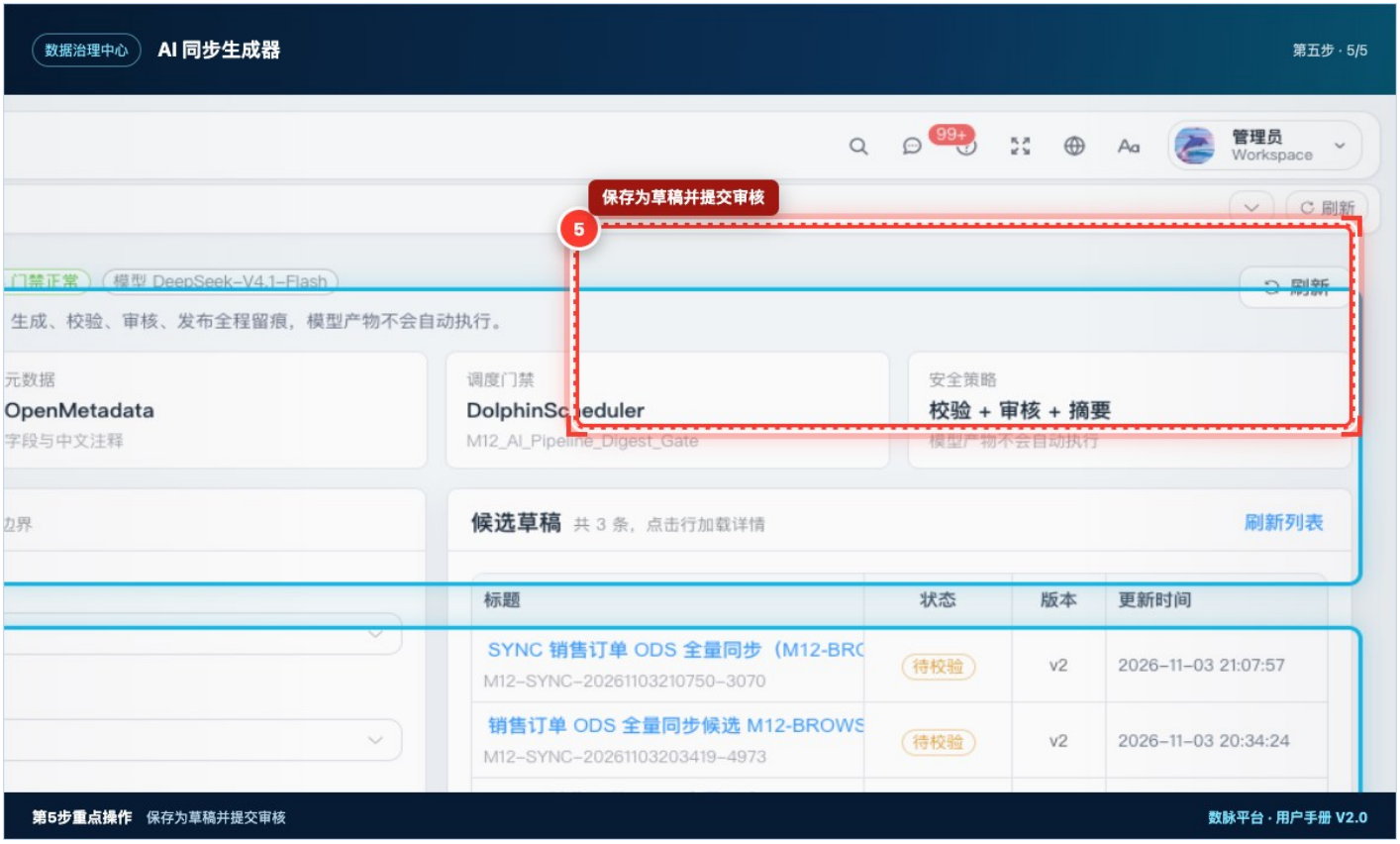
第四步：执行确定性校验


数据治理中心
AI 同步生成器
第四步 · 4/5
自然语言输入与生成
执行确定性校验
候选草稿列表
4
大模型
DeepSeek-V4.1-Flash
元数据
OpenMetadata
调度门禁
DolphinScheduler
安全策略
校验 + 审核 + 摘要
自然语言生成
候选草稿
共 3 条，点击行加载详情
标题
状态
版本
更新时间
SYNC 销售订单 ODS 全量同步 M12-BRO
M12-SYNC-20261103210750-3070
待校验
v2
销售订单 ODS 全量同步候选 M12-BROWS
M12-SYNC-20261103203419-4973
待校验
v2
SYNC 销售订单 ODS 全量同步 M12-BROV
M12-SYNC-20261103202837-3427
已审核
v1
白名单源表
方案标题
业务描述
AI 生成候选草稿
生成结果不会自动执行，必须先校验并人工审核。
bi_ods.ods_sale_order
OpenMetadata
ODS-销售订单原始表
字段
类型
中文注释
order_id
bigint
订单ID
order_no
varchar(64)
订单编号
order_date
date
下单日期
第4步重点操作 执行确定性校验
数脉平台 · 用户手册 V2.0

□□□ *40-04 □ 4*□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□



步骤图 40-05 第5步：保存为草稿并提交审核。红框标出保存为草稿并提交审核。
保存为草稿并提交审核；审核发布后才交给调度执行。

最终结果与效果
生成符合固定模板的同步草稿，包含字段映射、批次和安全配置；不会自动执行或改写生产数据。

操作提示： AI 草稿必须经过校验和人工审核；不要用自然语言要求绕过白名单。 “

常见问题
生成失败或超时。 检查 AI 服务状态、描述长度和源表字段；不要重复提交十几次，先看页面错误信息。选择不到源表。 表不在白名单或元数据未采集。
修改 JSON 后不能发布。 修改会重置校验和审核，必须重新校验、重新审核。

开发人员参考
接口： /bi/ai-pipeline/status、 /options?kind=SYNC、 /generate、 /drafts、 /drafts/{id}/validate|review|publish
权限： bi:etl:ai-sync
大模型只能选择模板和字段，不能生成任意连接串或执行任意组件。
发布只登记固定摘要门禁，不直接执行模型产物；内容 SHA-256 与审核记录绑定。

AI 加工设计器

功能定位
AI 加工设计器把业务描述转换成 Hop 转换候选流程和依赖 DAG
，例如过滤取消订单、补齐部门信息、统一金额单位、按订单日期排序后写入 DWD 。服务端固定可用组件、字段范围和写入边界， AI 不直接获得数据库写权限。

图 7-7 Hop 加工候选草稿与固定安全边界

适用角色
数据开发人员、数仓工程师和 ETL 管理员。

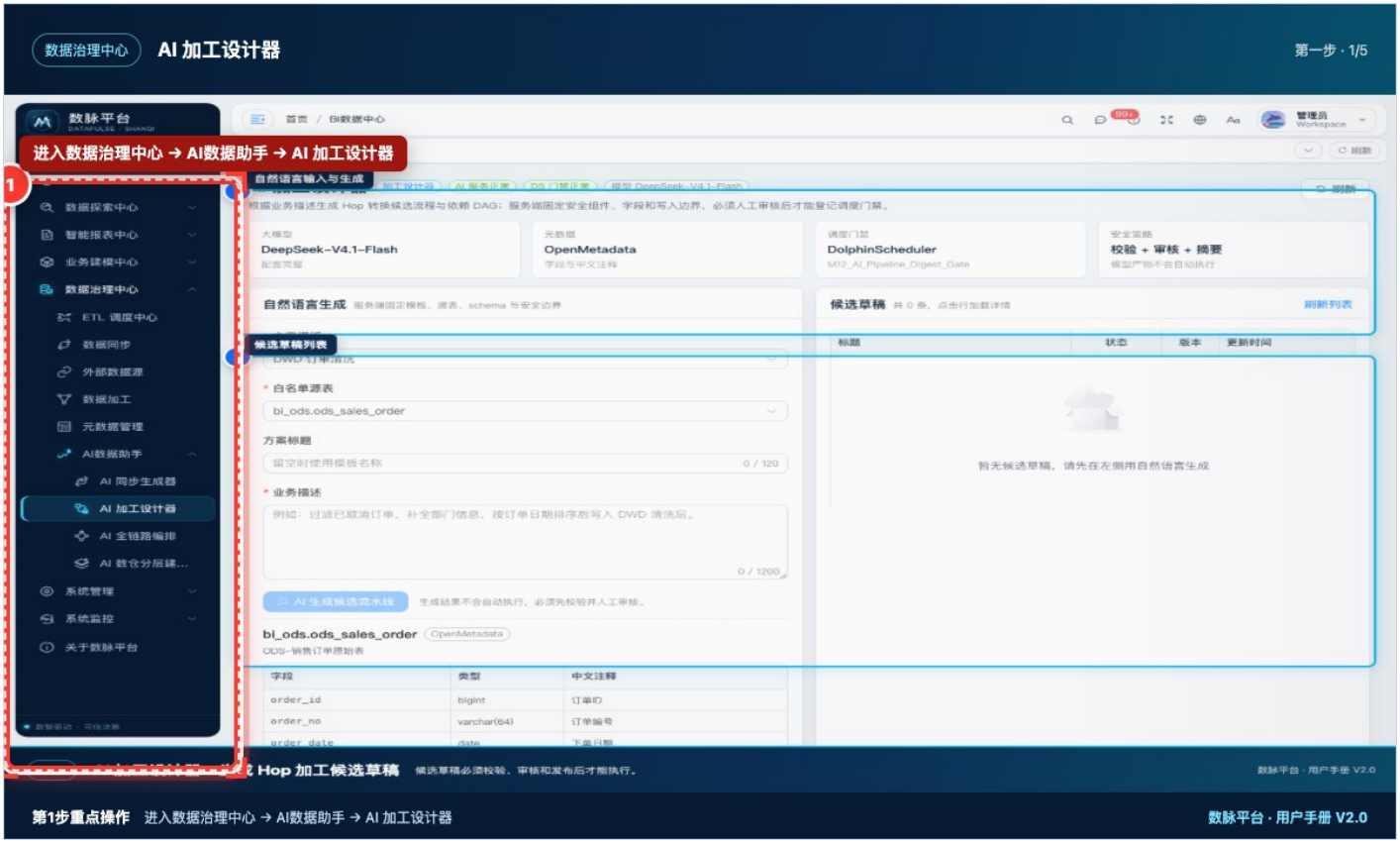
页面入口
数据治理中心 → AI 数据助手 → AI 加工设计器（ /bi/etl/ai/hop）。

操作步骤
1. 选择加工模板，例如 “DWD 订单清洗 ”。
2. 选择白名单 ODS 源表，查看字段与中文注释。
3. 描述清洗目标：保留哪些字段、过滤什么状态、如何补维度、写入哪张 DWD 表。
4. 生成候选流水线，在结构化预览中检查输入、转换、输出和字段映射。
5. 在 “DAG 编排 ”查看节点和依赖边，确认没有孤立节点、循环依赖或错误方向。
6. 执行确定性校验，修正字段类型、空值、主键和写入范围问题。
7. 提交人工审核并登记发布；发布后由调度执行，不在页面直接运行模型产物。

输入与输出
输入是 ODS 源表、业务描述和加工模板；输出是 Hop 候选流程、字段映射、 DAG
、校验结果、审核版本和调度门禁。

宏澜家电示例
输入 “过滤已取消订单，补客户和产品维度，金额统一为元，按订单日期写入 dwd_sales_order_clean”。AI 生成过滤、查找维度、字段重命名和写入节点；开发人员发现退货订单不应直接删除后，增加状态字段并调整
过滤条件，再重新校验审核。

客户应用演示：让 AI 起草 ODS→DWD 清洗逻辑
业务情境： 订单数据需要清洗取消订单、重复订单和空客户后再进入 DWD。

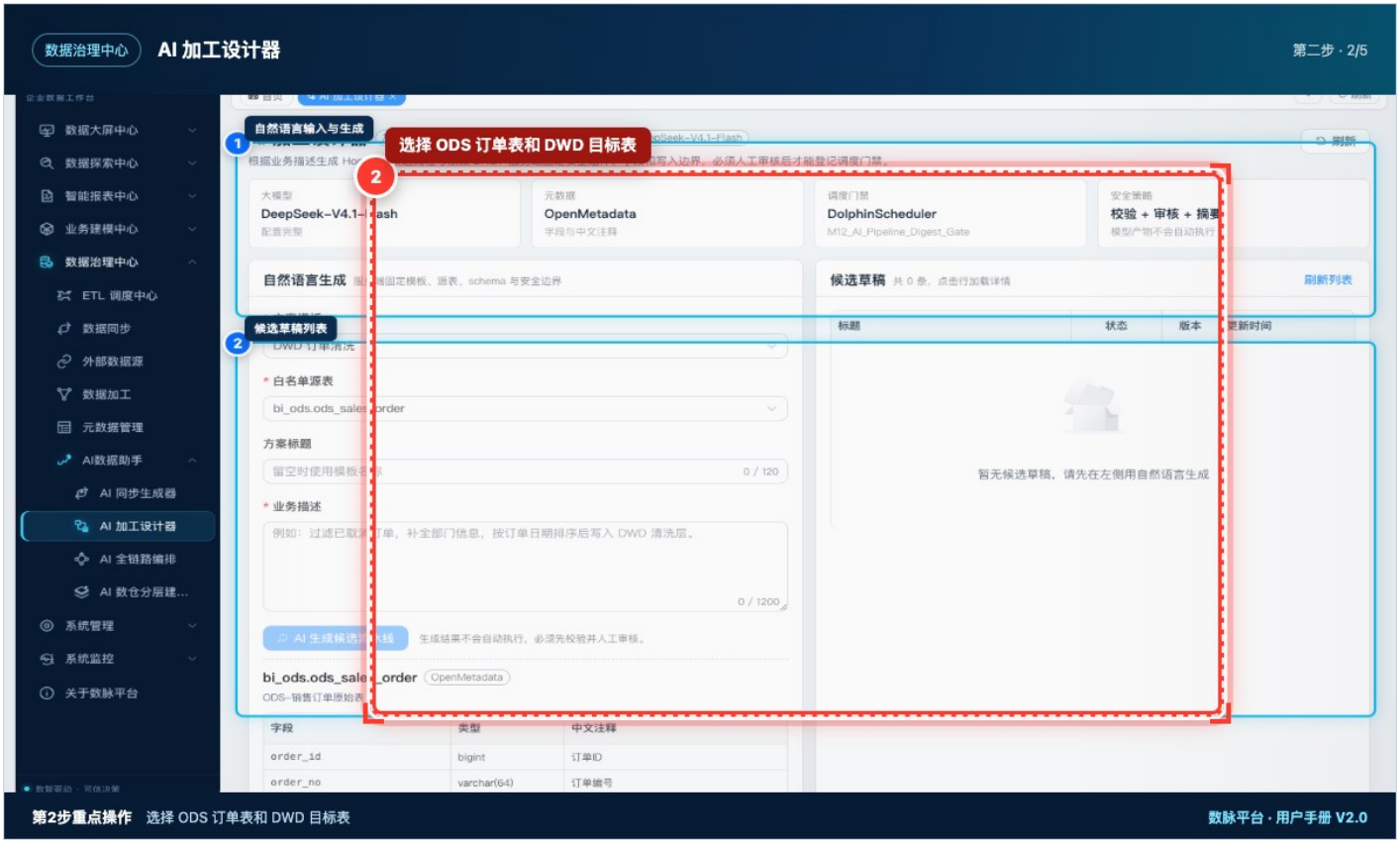


第一步：进入 “数据治理中心 → AI数据助手 → AI 加工设计器”

步骤图 *41-01* 第 *1*步：进入 “数据治理中心 → *AI*数据助手 → *AI* 加工设计器”。红框标出进入数据治理中心 → *AI* 数据助手 → *AI* 加工设计器。

进入 “数据治理中心 → AI数据助手 → AI 加工设计器”。

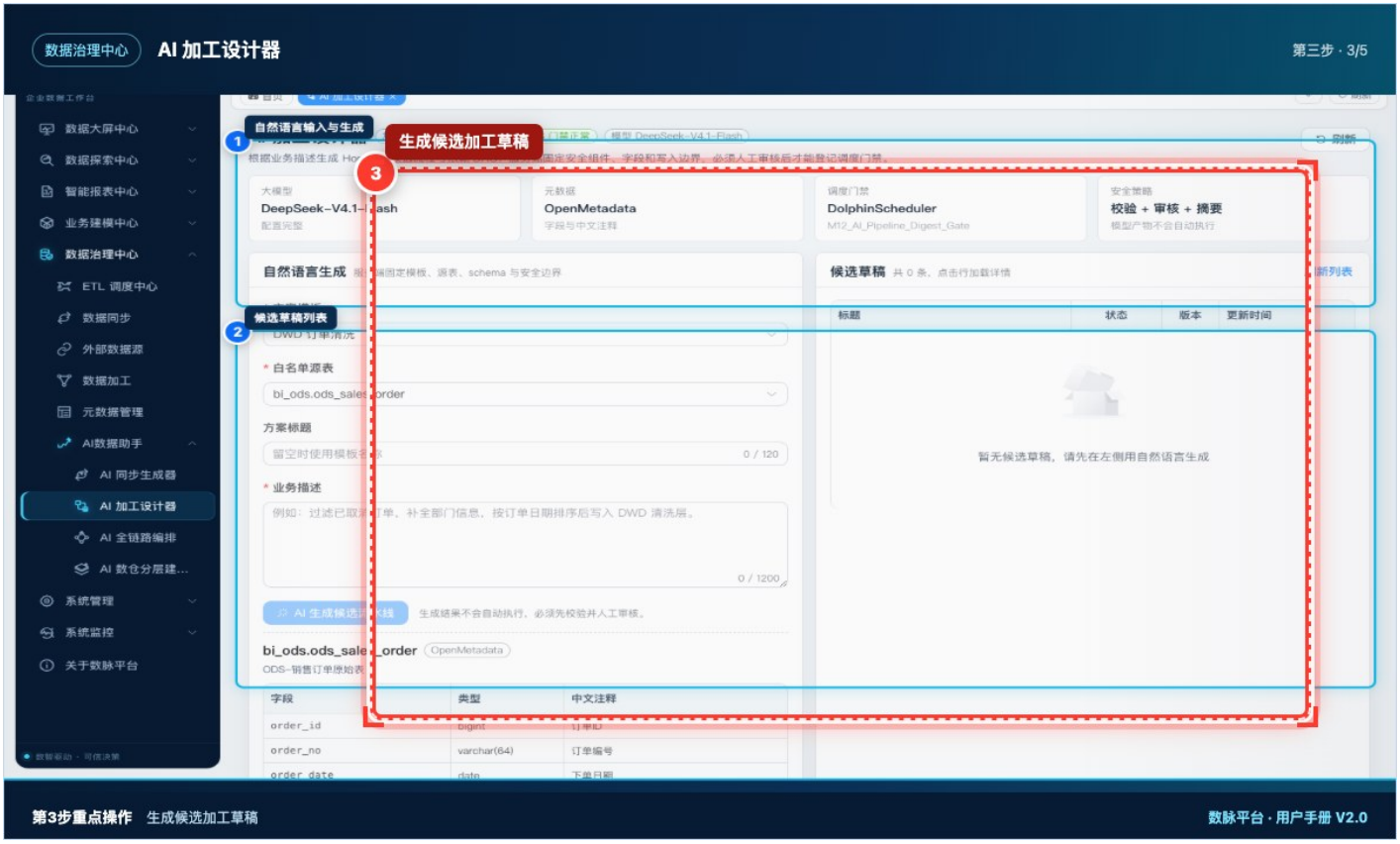
第二步：选择 ODS 订单表和 DWD 目标表



步骤图 41-02 第2步：选择 ODS 订单表和 DWD 目标表。红框标出选择 ODS 订单表和 DWD 目标表。

选择 ODS 订单表和 DWD 目标表，描述清洗规则与输出字段。

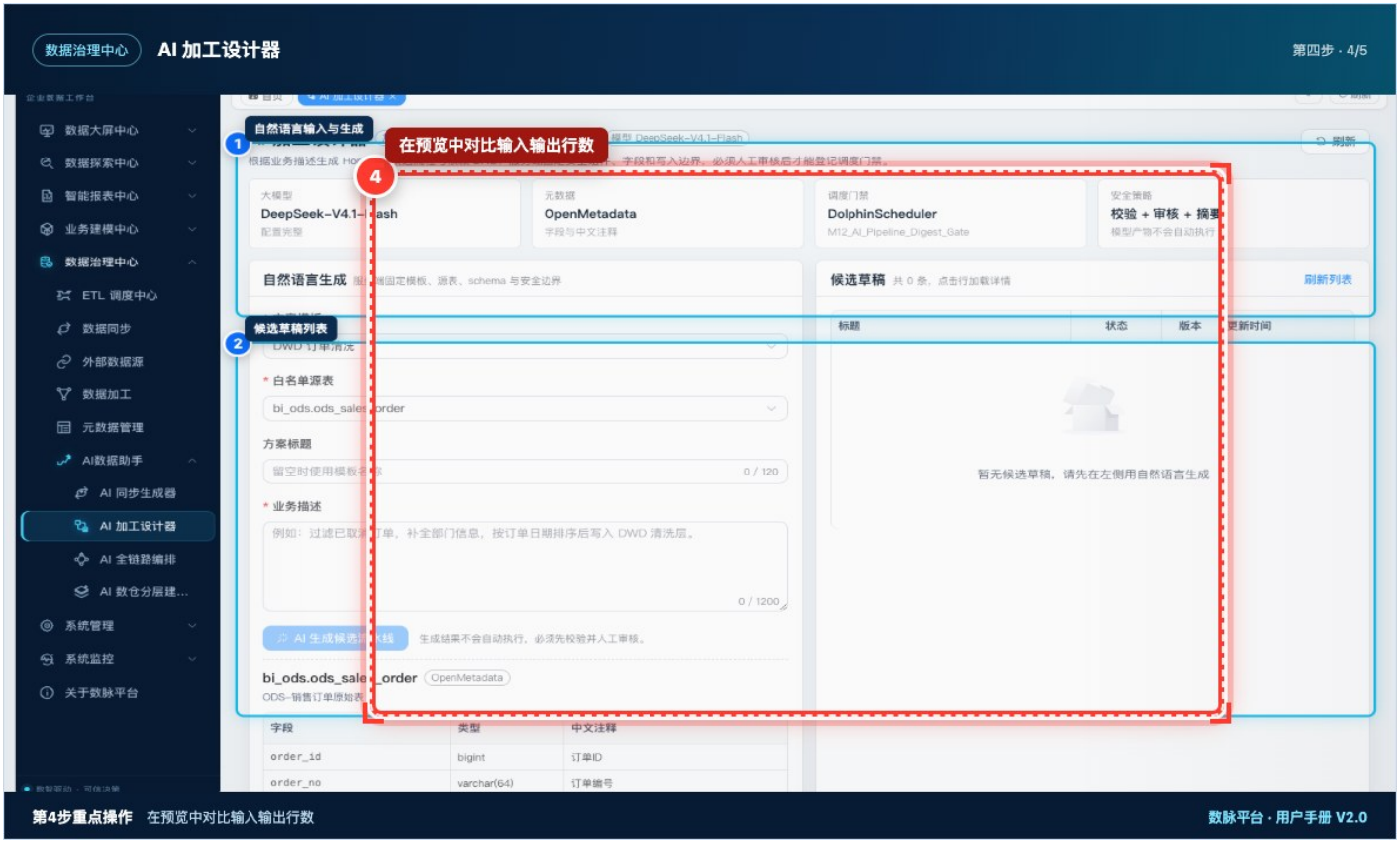
第三步：生成候选加工草稿


数据治理中心
AI 加工设计器
第三步 · 3/5
自然语言输入与生成
生成候选加工草稿
候选草稿列表
数据大屏中心
数据探索中心
智能报表中心
业务建模中心
数据治理中心
ETL 调度中心
数据同步
外部数据源
数据加工
元数据管理
AI数据助手
AI 同步生成器
AI 加工设计器
AI 全链路编排
系统管理
系统监控
关于数脉平台
大模型
DeepSeek-V4.1-Flash
元数据
OpenMetadata
调度门禁
DolphinScheduler
安全策略
校验 + 审核 + 摘要
自然语言生成
候选草稿
暂无候选草稿，请先在左侧用自然语言生成
方案标题
业务描述
bi_ods.ods_sales_order
字段
类型
中文注释
order_id
order_no
varchar(64)
订单编号
order_date
date
第3步重点操作 生成候选加工草稿
数脉平台 · 用户手册 V2.0

□□□ *41-03 □ 3*□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□



步骤图 41-04 第4步：在预览中对比输入输出行数。红框标出在预览中对比输入输出行数。

在预览中对比输入输出行数，检查空值和异常金额。

第五步：保存、校验、审核并发布



步骤图 41-05 第5步：保存、校验、审核并发布。红框标出保存、校验、审核并发布。
保存、校验、审核并发布；执行仍由调度中心负责。

最终结果与效果
得到结构化加工草稿和可核对样例，清洗逻辑可追溯，发布前不会影响生产数据。

操作提示： 业务规则要写清楚，例如 “取消订单不计入销售额 ”，不要只写 “清洗数据 ”。 “

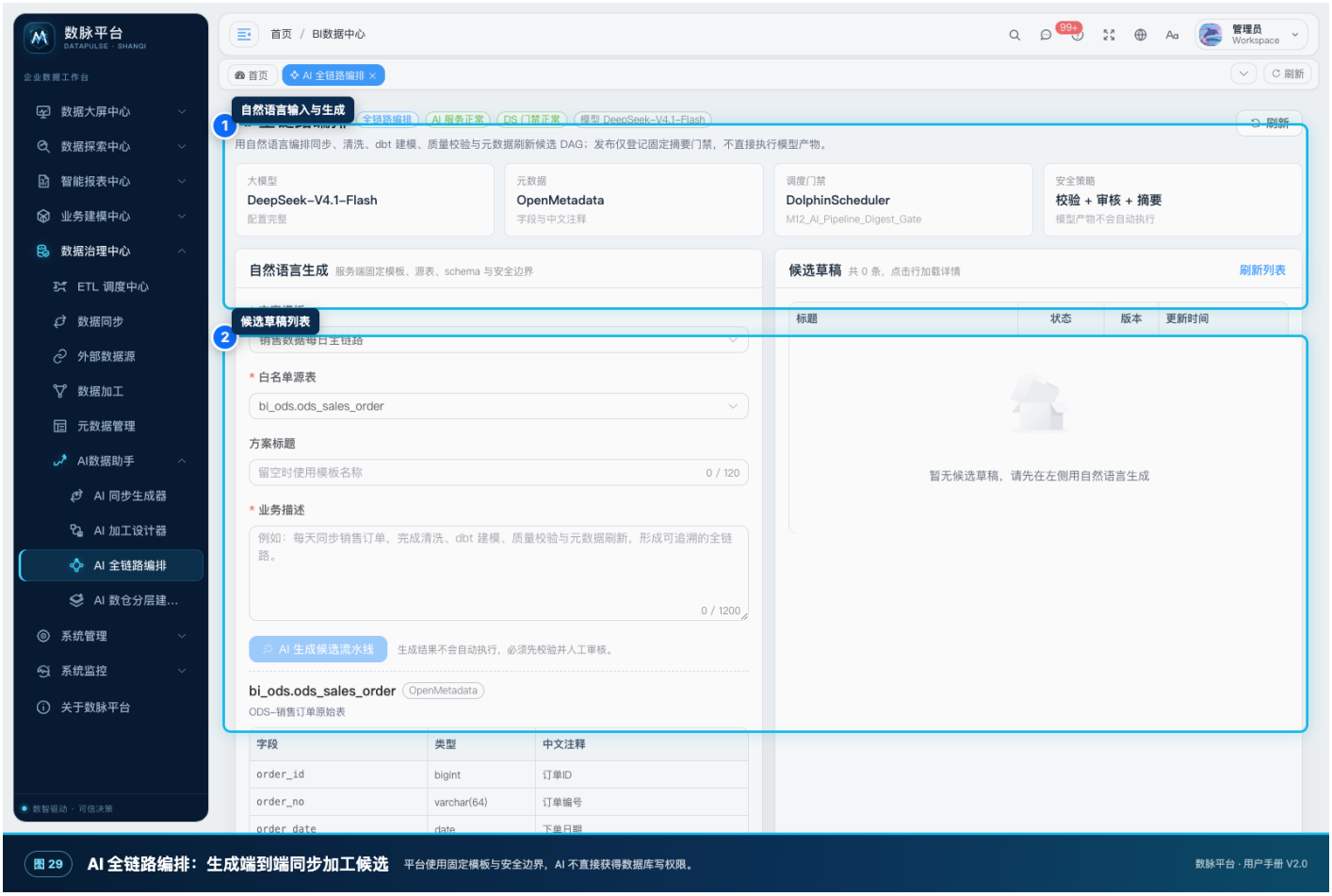
常见问题
候选 DAG 缺节点。 业务描述可能过于笼统；补充输入表、过滤条件、关联维度和目标表。校验提示字段越界。 AI 只能使用白名单字段；先在元数据中登记字段或调整方案。审核通过后还能改吗。 修改会生成新版本并使原审核失效，必须重新走校验和审核。

开发人员参考
接口： /bi/ai-pipeline/options?kind=HOP、 /generate、 /drafts/{id}/validate|review|publish、 /versions
权限： bi:etl:ai-hop
组件、字段、源表和目标表由服务端模板约束， JSON 仅允许编辑白名单结构化字段。
审核绑定 SHA-256 摘要，发布只登记 DolphinScheduler 门禁，不自动执行 Hop 产物。

AI □□□□

□□□

AI □□□□□□□ “□□□□□ ”□□□□□□□□□□□□□□□□□□□□□□ dbt □□□□□□□□□□□□□□□ DAG □□



□ *7-8* □□□□□□□□□□□□□□□

□□□

□□□□□□ ETL □□□□□□□□□□□□□□□□□□□□□□

页面入口
数据治理中心 → AI 数据助手 → AI 全链路编排（ /bi/etl/ai/whale）。

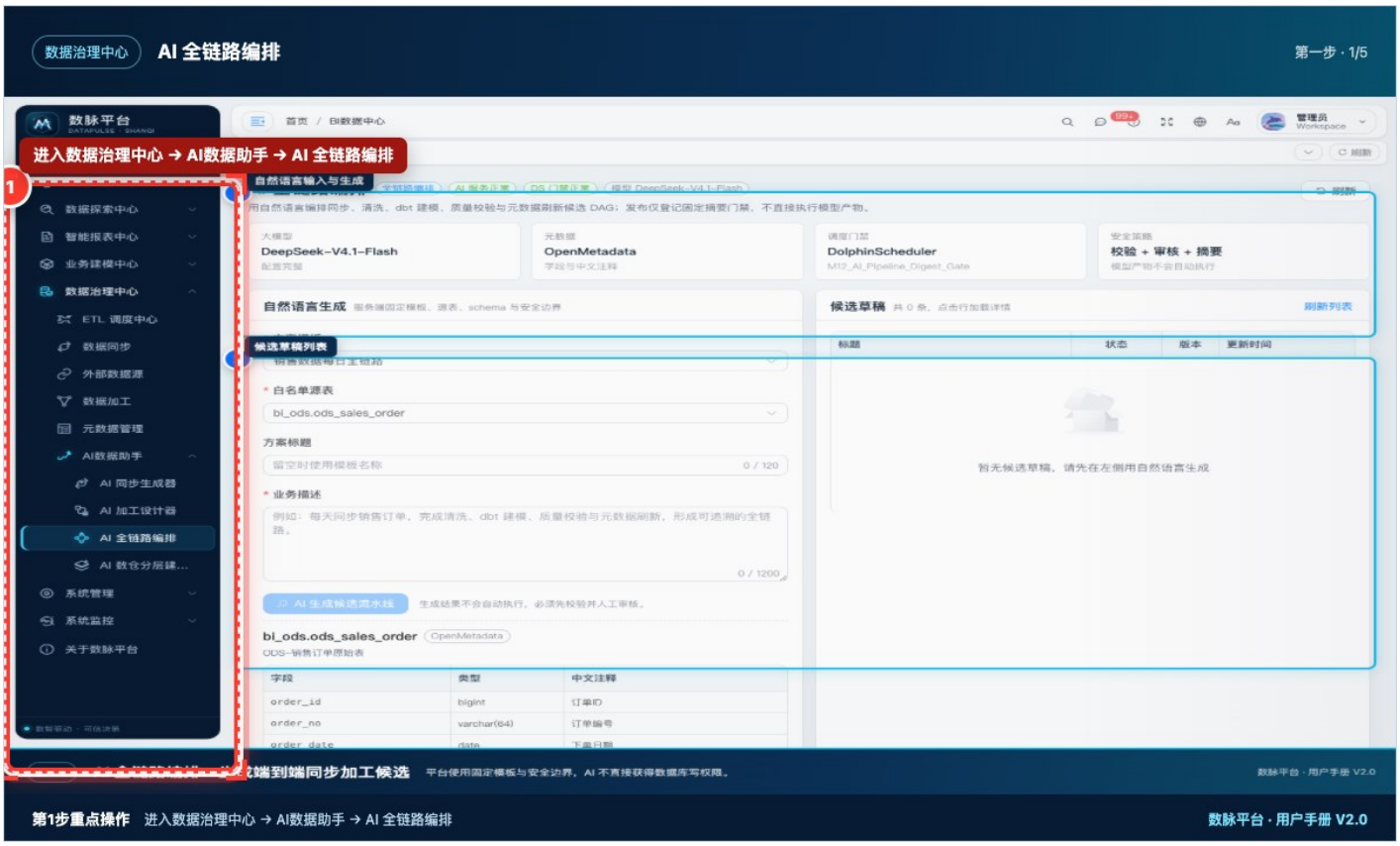
操作步骤
1. 选择全链路模板，例如 “销售数据每日主链路 ”。
2. 选择白名单源表，填写业务描述：同步频率、目标分层、清洗规则、质量要求和刷新范围。
3. 生成候选 DAG，查看节点类型、目标表、依赖边和关键参数。
4. 检查链路是否包含 “源表就绪 → 同步 → DWD 清洗 → DWS 汇总 → 质量测试 → 元数据刷新 ” 的必要阶段。
5. 对每个节点检查失败策略、重试次数、超时和依赖条件，避免一个节点失败后下游仍继续。
6. 执行确定性校验，处理循环依赖、缺失字段、越权表和非白名单组件。
7. 提交审核并登记发布；发布后仍需在调度中心按正常流程执行和监控。

输入与输出
输入是源表、业务描述、分层目标和模板；输出是端到端 DAG
、节点参数、依赖关系、校验结果、审核版本和调度门禁。

宏澜家电示例
输入 “
每天同步销售订单，清洗后按区域品类汇总，执行金额非负和主键唯一测试，刷新元数据，供经营看板使用 ”。
AI 生成五段式 DAG；数据架构师补充 “质量失败阻断下游 ”的条件后审核发布，避免脏数据进入 ADS。

客户应用演示：生成 “

销售数据每日主链路 ”
业务情境： 需要把同步、清洗、汇总、质量和刷新串成一条每日 DAG。

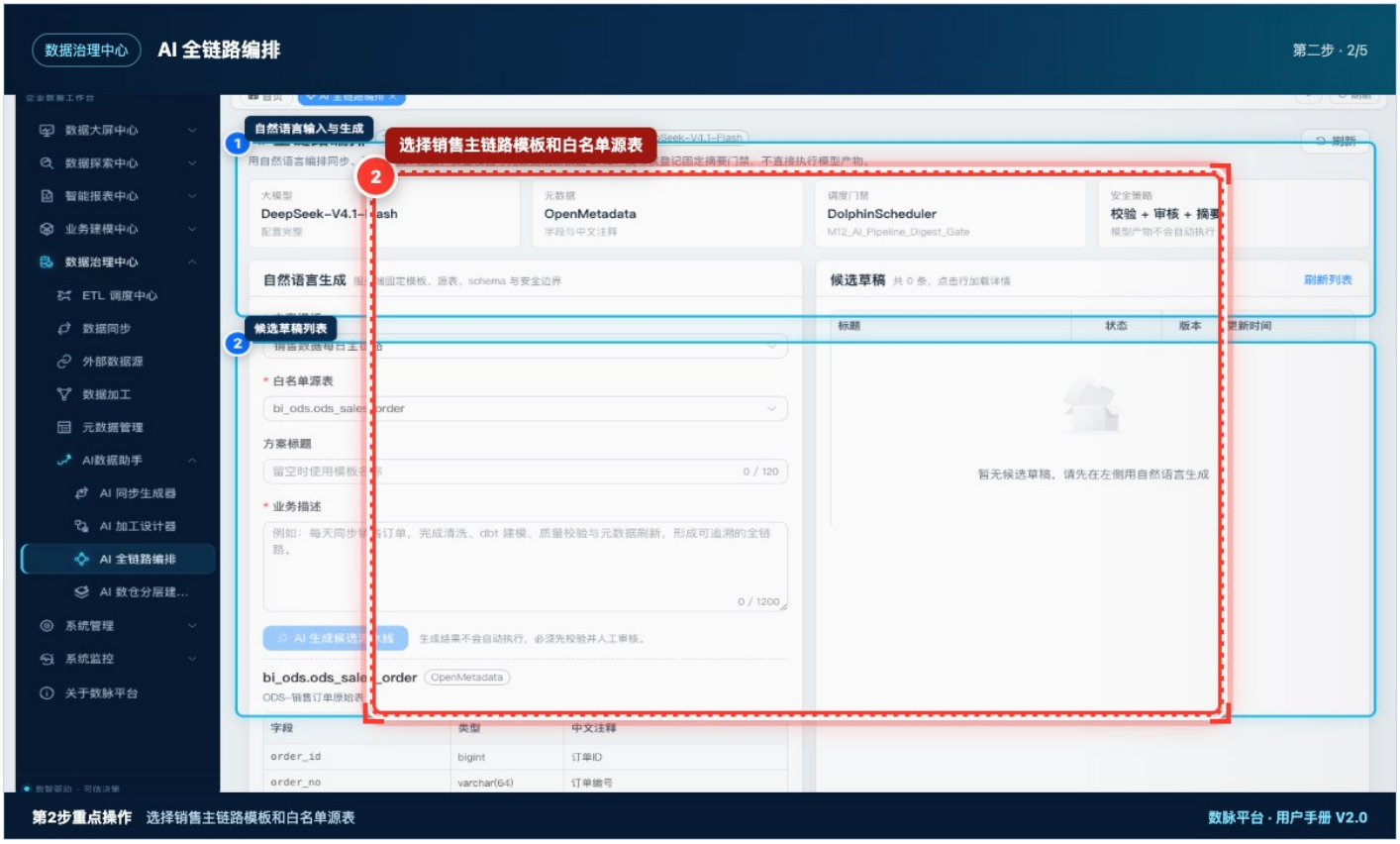


第一步：进入 “数据治理中心 → AI数据助手 → AI 全链路编排 ”

步骤图 *42-01* 第 *1*步：进入 “数据治理中心 → *AI*数据助手 → *AI* 全链路编排 ”。红框标出进入数据治理中心 → *AI* 数据助手 → *AI* 全链路编排。

进入 “数据治理中心 → AI数据助手 → AI 全链路编排 ”。

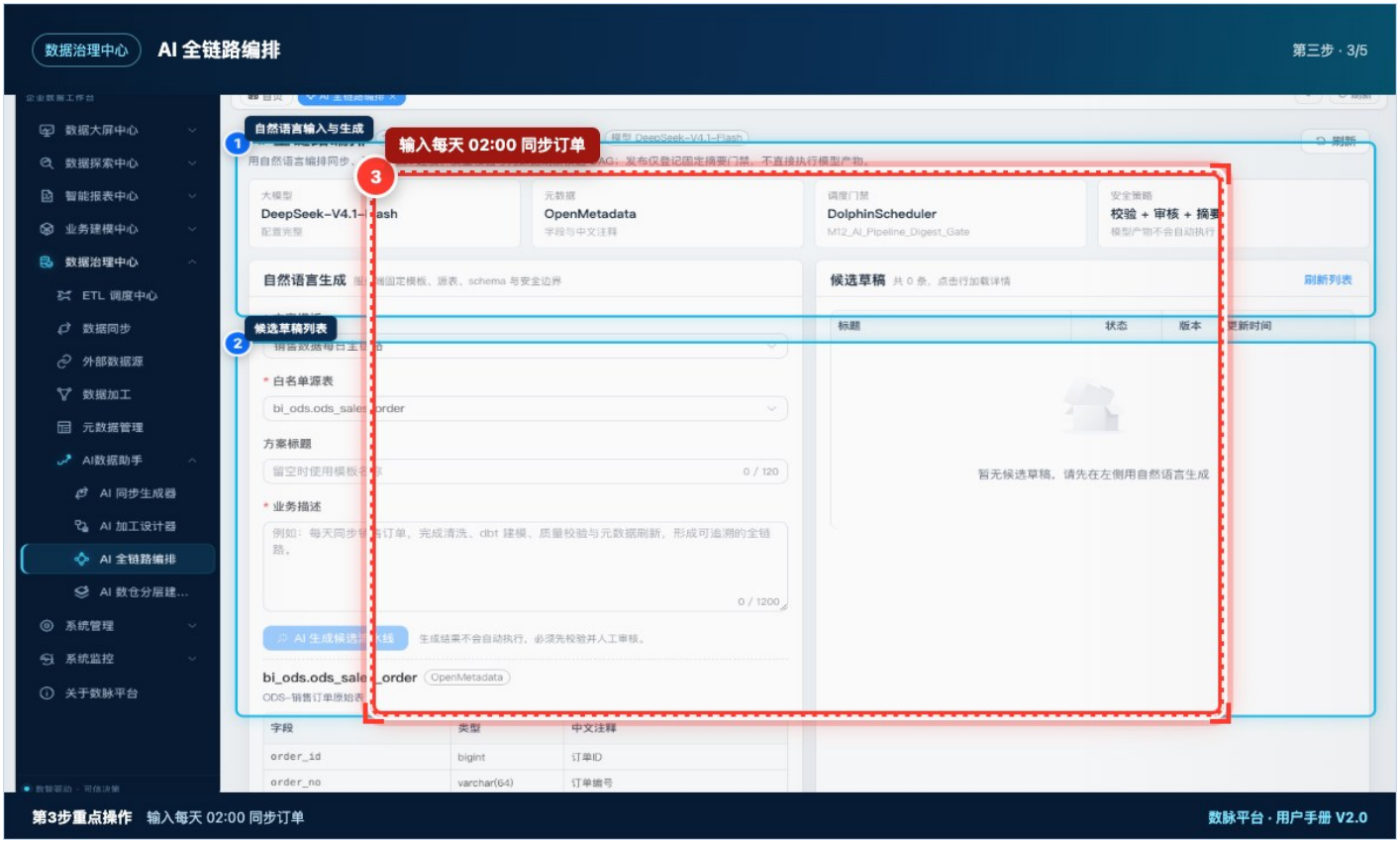
第二步：选择销售主链路模板和白名单源表



步骤图 42-02 第2步：选择销售主链路模板和白名单源表。红框标出选择销售主链路模板和白名单源表。

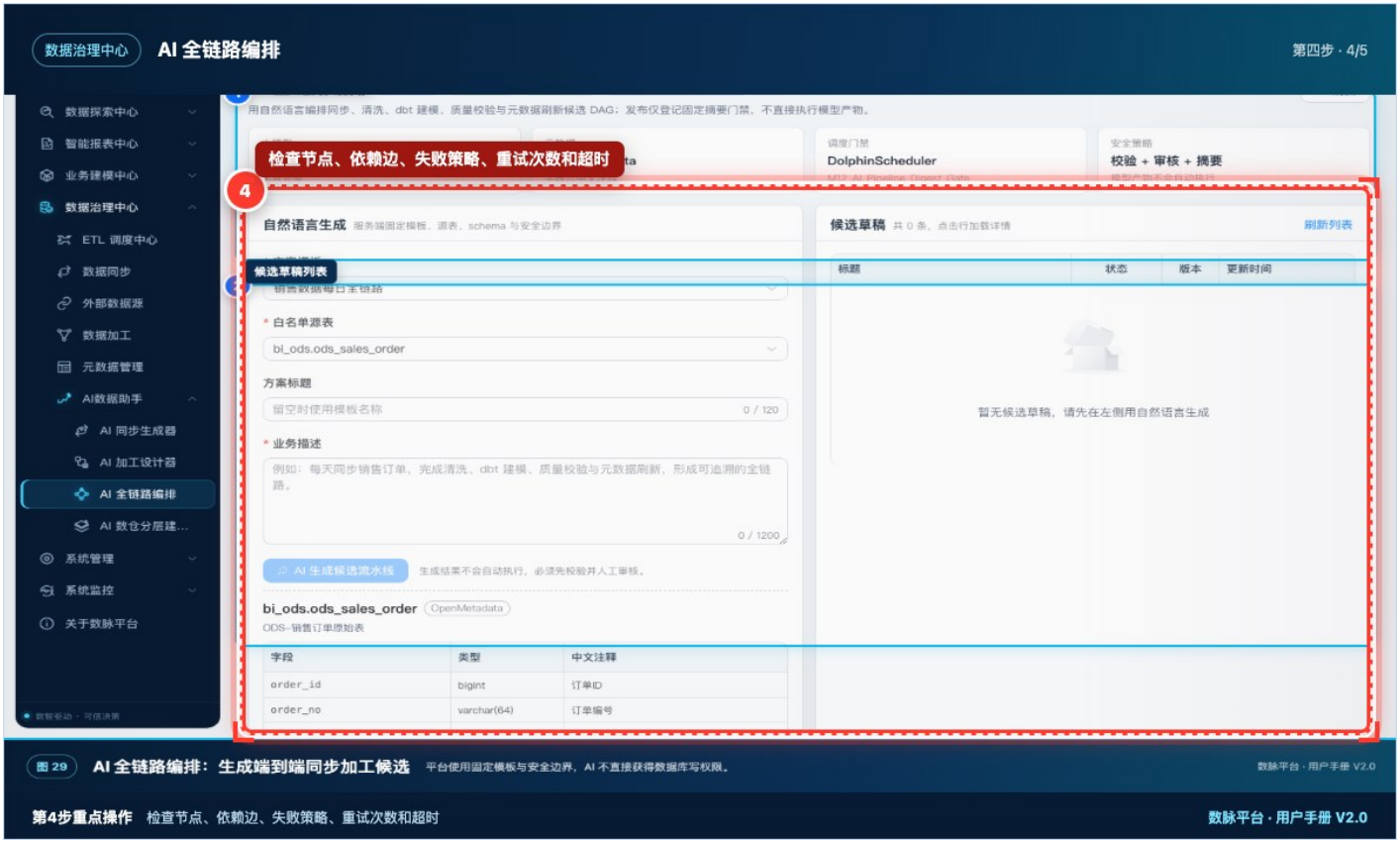
选择销售主链路模板和白名单源表。

第三步：输入 “每天 02:00 同步订单



步骤图 42-03 第3步：输入 “每天 02:00 同步订单。红框标出输入每天 02:00 同步订单。
输入 “每天 02:00 同步订单，清洗后汇总区域品类，执行金额非负和主键唯一测试，再刷新数据集 ”。

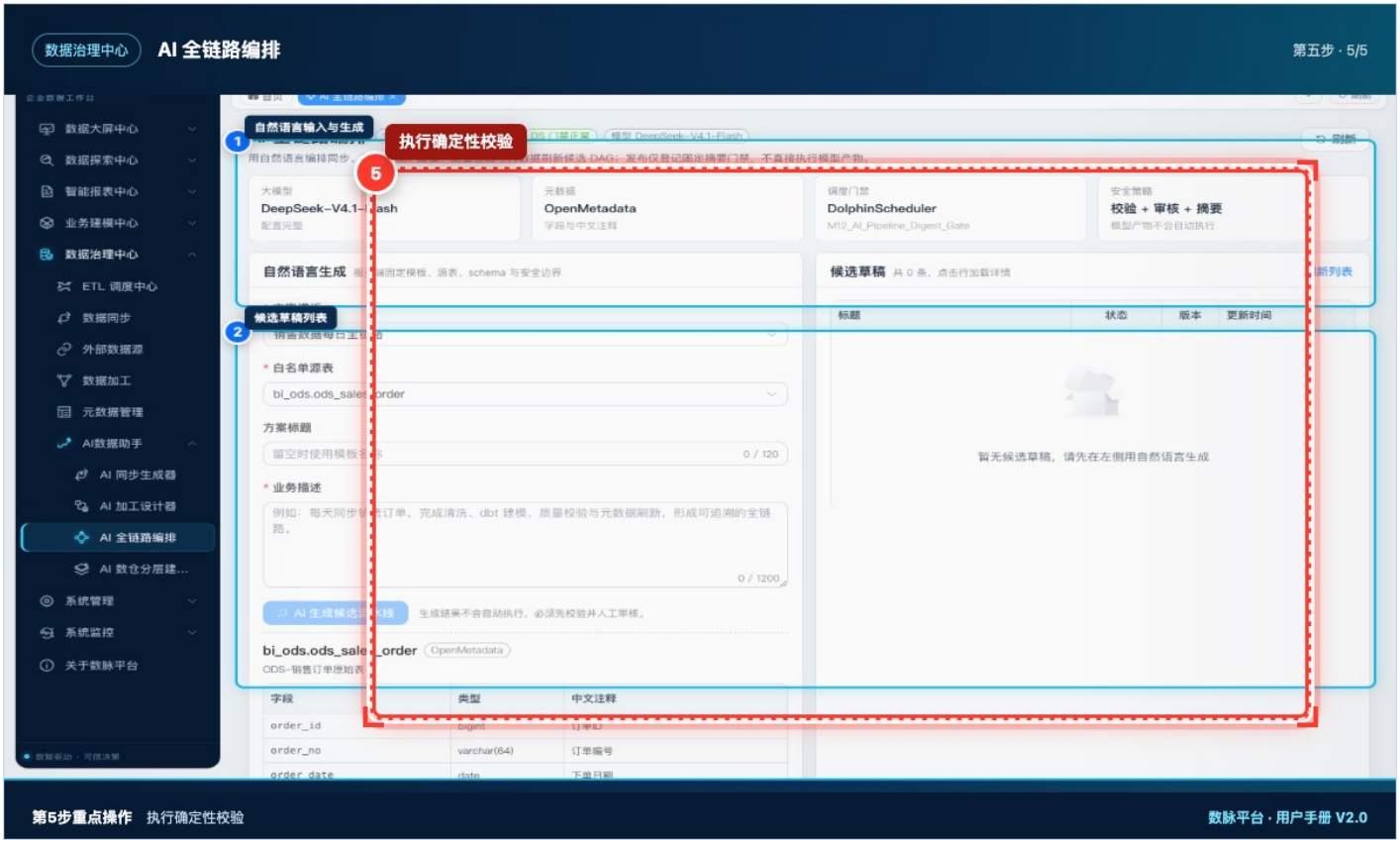
第四步：检查节点、依赖边、失败策略、重试
次数和超时



步骤图 42-04 第4
步：检查节点、依赖边、失败策略、重试次数和超时。红框标出检查节点、依赖边、失败策略、重试次数和超时。

检查节点、依赖边、失败策略、重试次数和超时。

第五步：执行确定性校验



步骤图 42-05 第5步：执行确定性校验。红框标出执行确定性校验。
执行确定性校验，补充 “质量失败阻断下游 ”，提交审核发布。

最终结果与效果
生成端到端 DAG 草稿，包含同步、加工、汇总、质量和刷新阶段；发布只登记调度门禁，不自动运行。

操作提示： 确保链条有源表就绪、失败阻断和批次追踪，避免半成品进入看板。 “

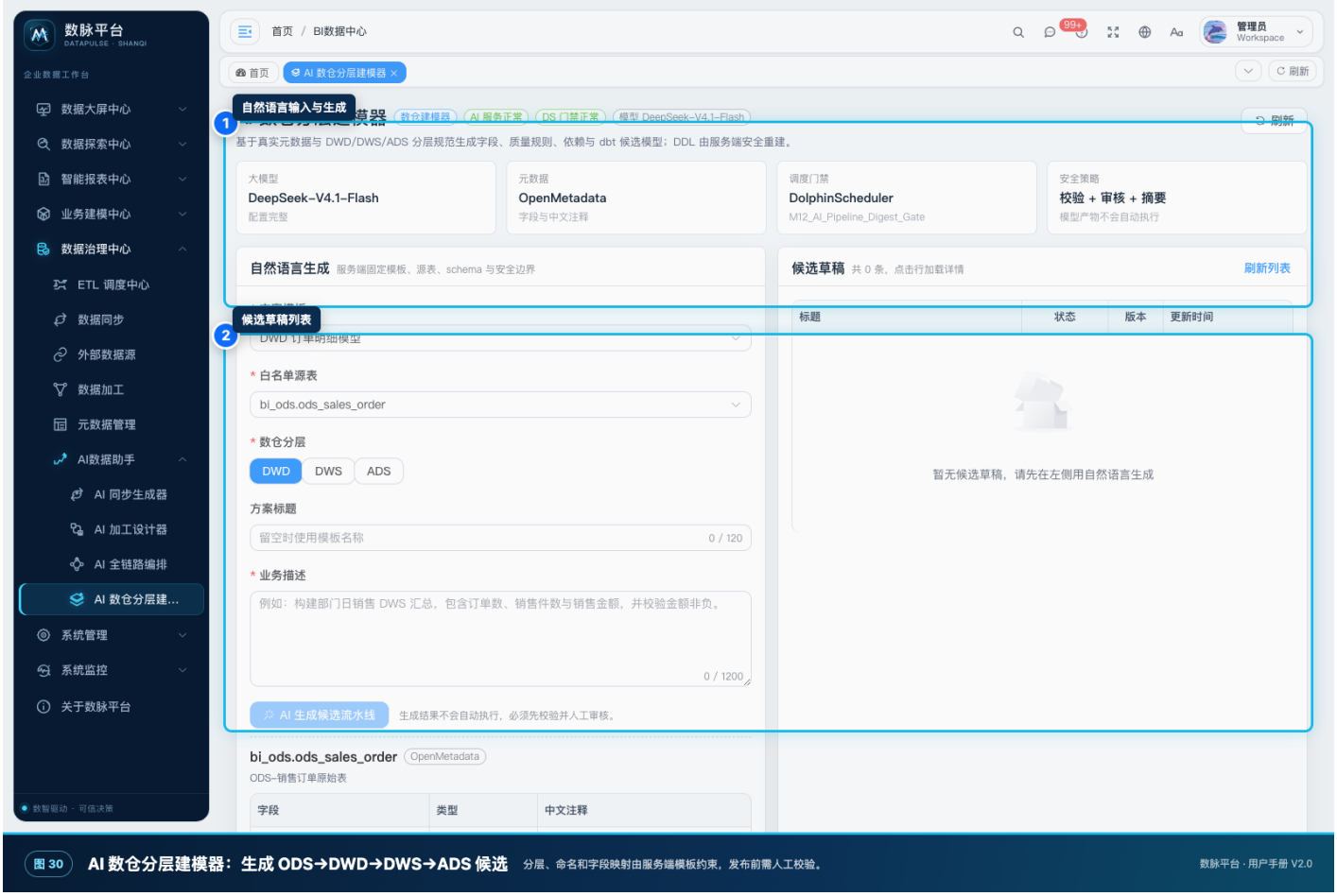
常见问题
DAG 看起来完整但不能发布。 先看校验结果，通常是字段越界、依赖循环或缺少目标表。
全链路是否会自动执行。 不会。发布只登记调度门禁和摘要，实际执行由 DolphinScheduler 负责。某个节点失败如何补数。 从失败节点开始按依赖顺序补跑，避免只补最后一层造成口径断层。

开发人员参考
接口： /bi/ai-pipeline/options?kind=WHALE、 /generate、 /drafts/{id}/validate|review|publish、 /versions/{id}/rollback
权限： bi:etl:ai-whale
服务端对 DAG 做节点白名单、依赖环检测、字段边界和摘要校验。
发布不直接调用执行器，避免 AI 生成内容未经人工确认就改变生产数据。

AI □□□□□□

□□□

AI □□□□□□□□□□□□□□□□ DWD/DWS/ADS □□□□□□□□□□□□□□□□□□□□□□□□
dbt □□□□□□□□□□□□□□□□□□□□□□□□ DDL
□□□□□□□□□□□□□□□□□□□□□□□□□□□□



□ *7-9 ODS → DWD → DWS → ADS □□□□□□□*

□□□

□□□□□□□□□□□□□ dbt □□□□□□□□□□□

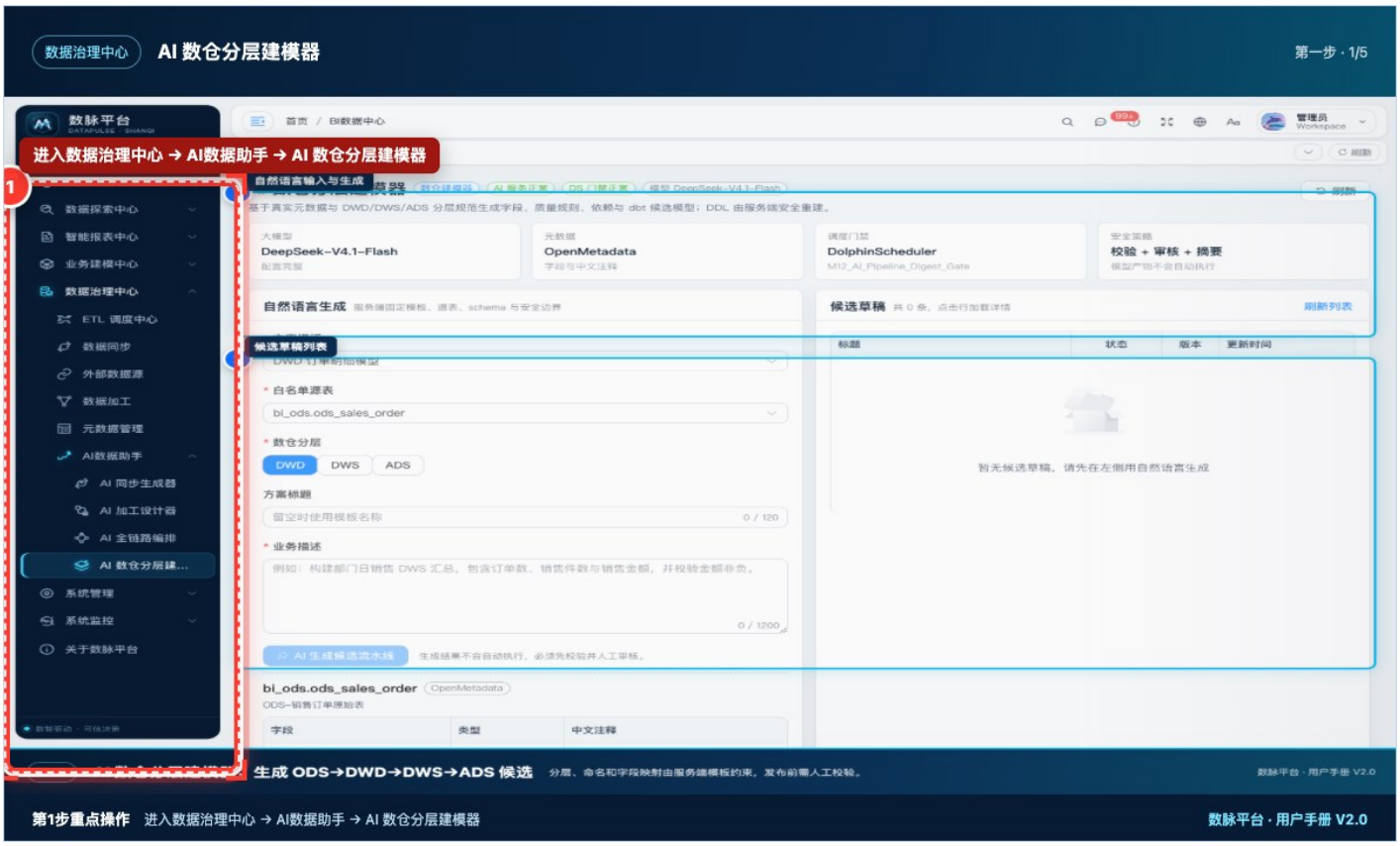
页面入口
数据治理中心 → AI 数据助手 → AI 数仓分层建模器（ /bi/etl/ai/warehouse）。

操作步骤
1. 选择分层模板和数仓分层： DWD、 DWS 或 ADS。
2. 选择白名单源表，确认字段、类型和注释已经采集。
3. 描述业务目标，例如 “构建部门日销售 DWS
汇总，包含订单数、销售件数与销售金额，并校验金额非负 ”。
4. 生成候选模型，查看字段、主键、分区、度量、质量规则和依赖。
5. 在 “结构化预览 ”检查服务端重建的 DDL；不要手工拼接建表语句。
6. 在 “DAG 编排 ”确认上游模型和下游依赖，避免跨层反向依赖。
7. 校验、审核、发布；发布后在模型设计和发布中心继续跟踪 dbt、 Doris、OpenMetadata 和
Superset 状态。

输入与输出
输入是源表元数据、目标分层、业务描述和模板；输出是字段定义、主键与分区、质量规则、依赖、 dbt 候选模型、 DDL 和审核版本。

宏澜家电示例
选择 DWS，描述 “
按部门、日期汇总销售订单，输出订单数、销售件数、销售金额、折扣额和毛利率，金额不得为负 ”。AI
生成汇总字段和质量规则；建模人员补充 “取消订单不计入订单数 ”
的口径后审核。发布后该模型成为部门经营看板的数据集来源。

客户应用演示：生成部门日销售 DWS 模型
业务情境： 经营看板需要按部门和日期汇总销售订单，并统一金额口径。

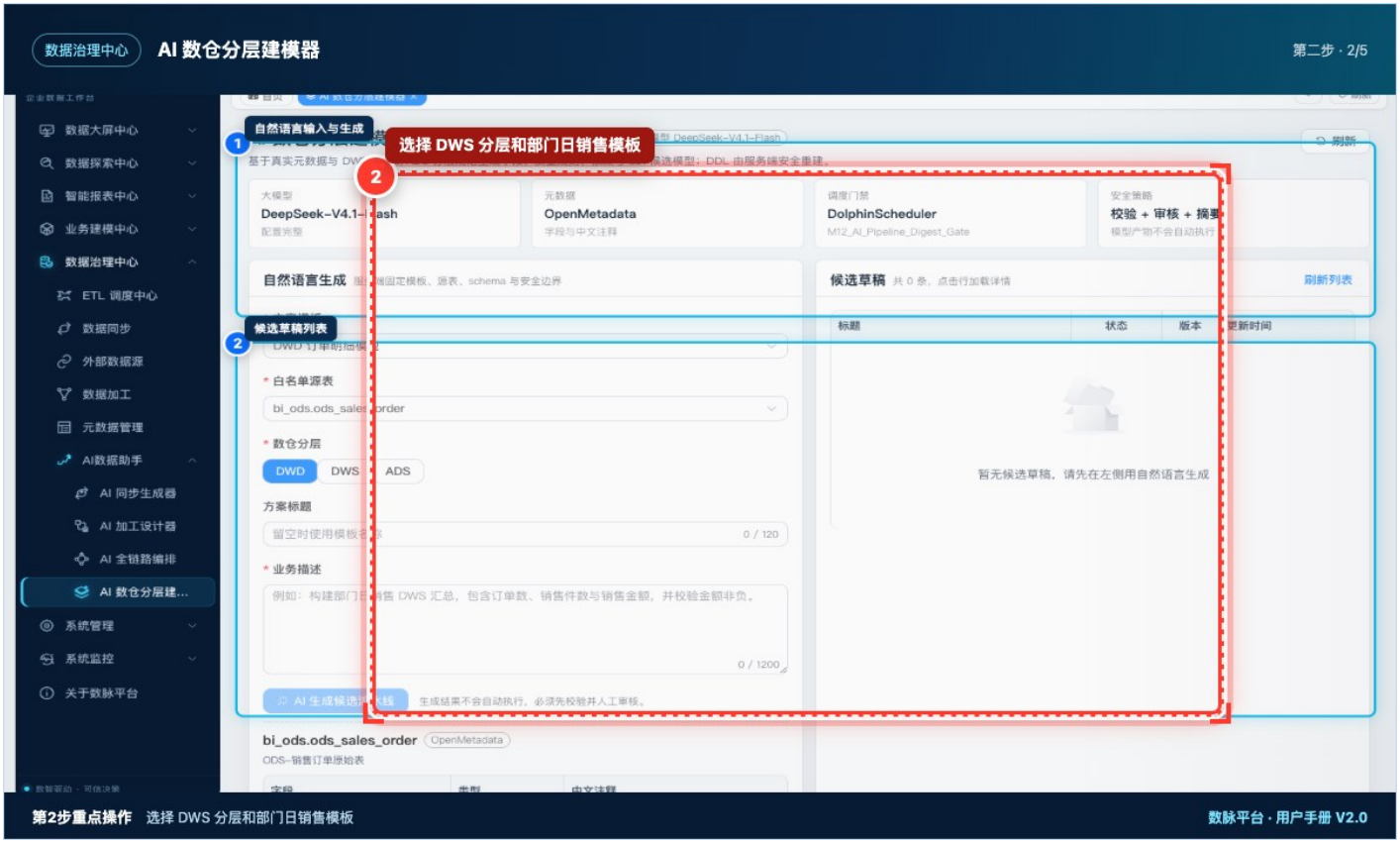


第一步：进入“数据治理中心 → AI数据助手 → AI 数仓分层建模器”

步骤图 *43-01* 第 *1*步：进入 “数据治理中心 → *AI*数据助手 → *AI* 数仓分层建模器 ”。红框标出进入数据治理中心 → *AI*数据助手 → *AI* 数仓分层建模器。

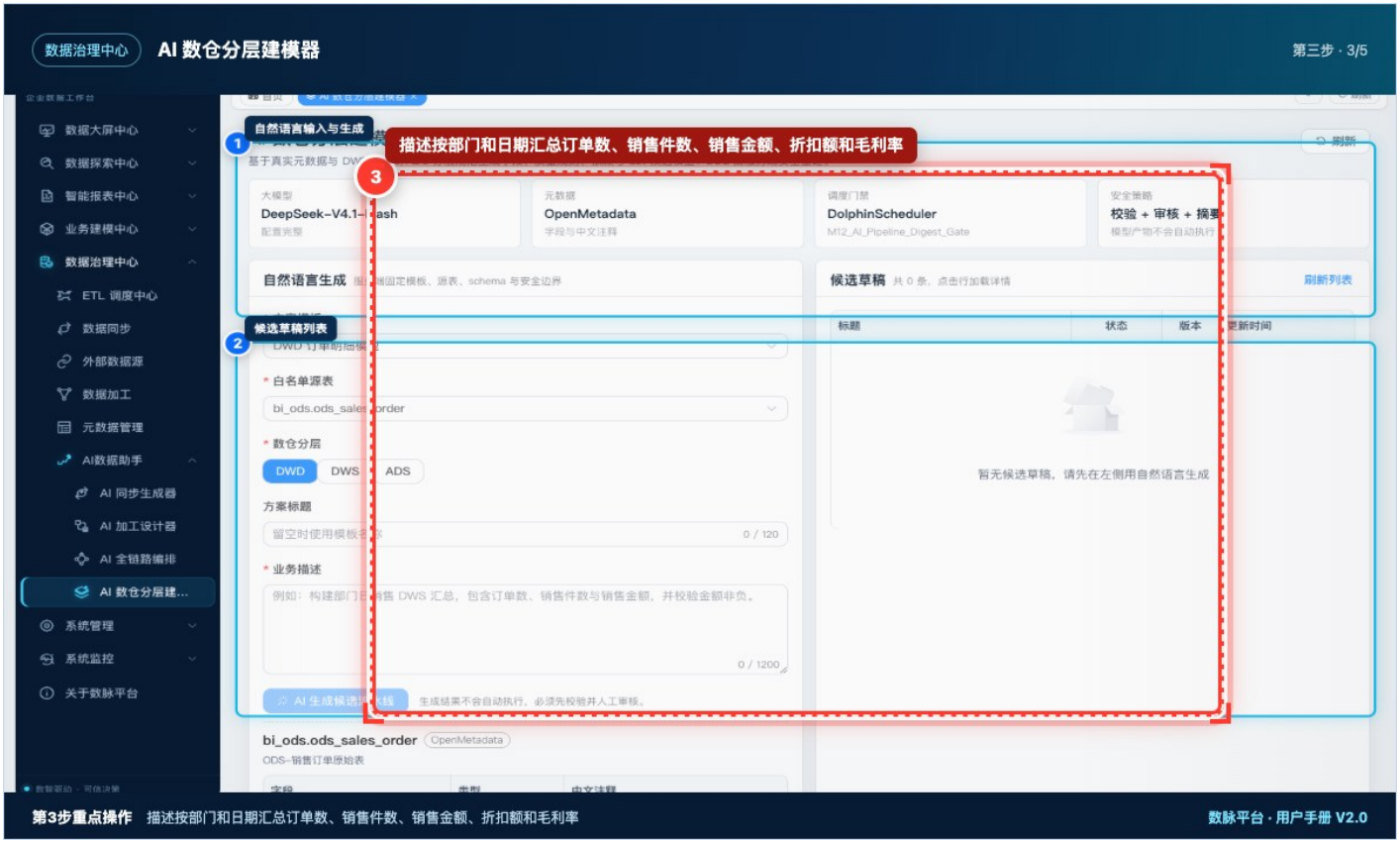
进入 “数据治理中心 → AI数据助手 → AI 数仓分层建模器 ”。

第二步：选择 DWS 分层和部门日销售模板



步骤图 43-02 第2步：选择 DWS 分层和部门日销售模板。红框标出选择 DWS 分层和部门日销售模板。选择 DWS 分层和部门日销售模板，选择白名单源表。

第三步：描述 “
按部门和日期汇总订单数、销售件数、销售金
额、折扣额和毛利率

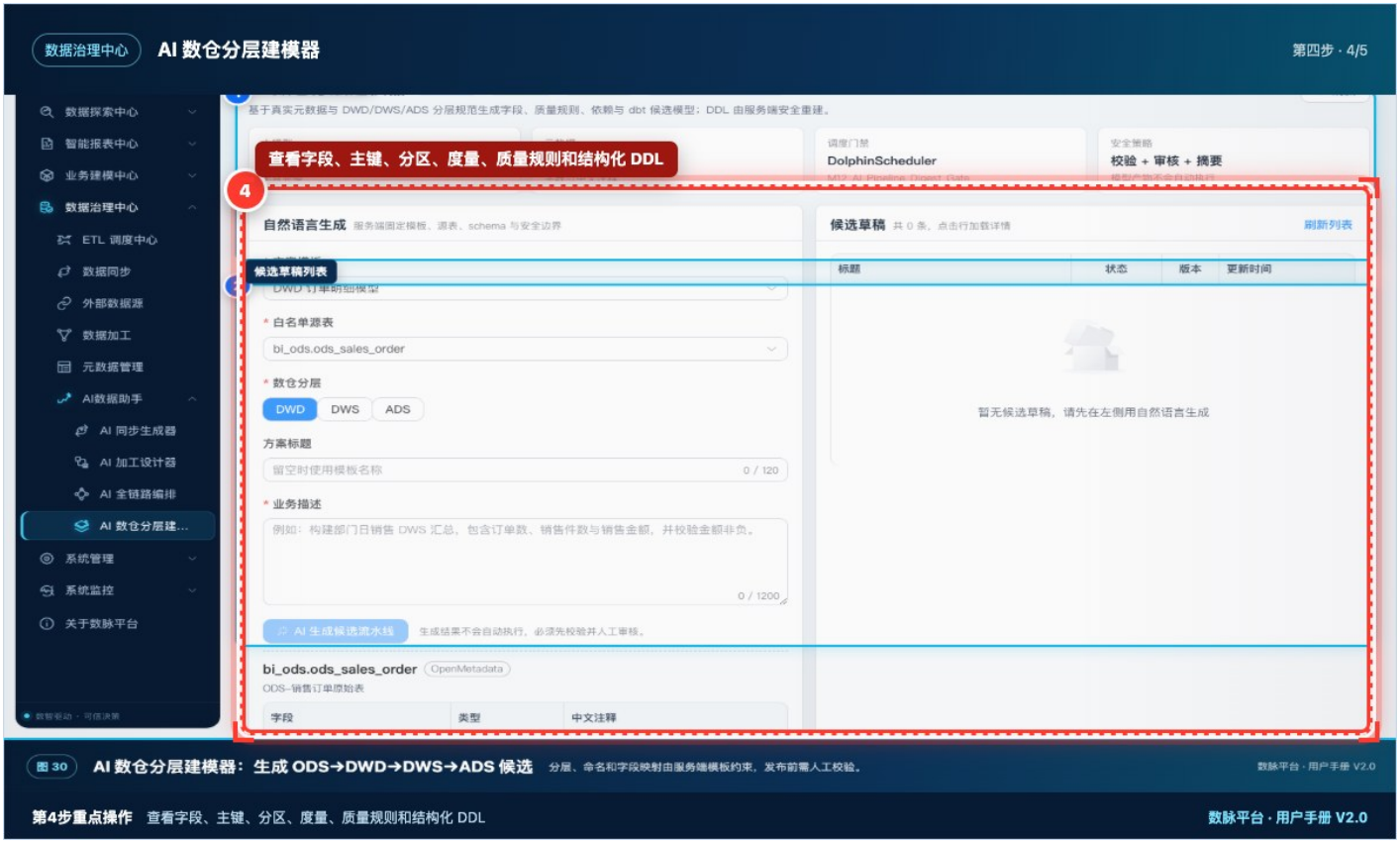


步骤图 43-03 第3步：描述 “
按部门和日期汇总订单数、销售件数、销售金额、折扣额和毛利率。红框标出描述按部门和日期汇总订单数、
销售件数、销售金额、折扣额和毛利率。

描述 “按部门和日期汇总订单数、销售件数、销售金额、折扣额和毛利率，金额不得为负 ”。

第四步：查看字段、主键、分区、度量、质量

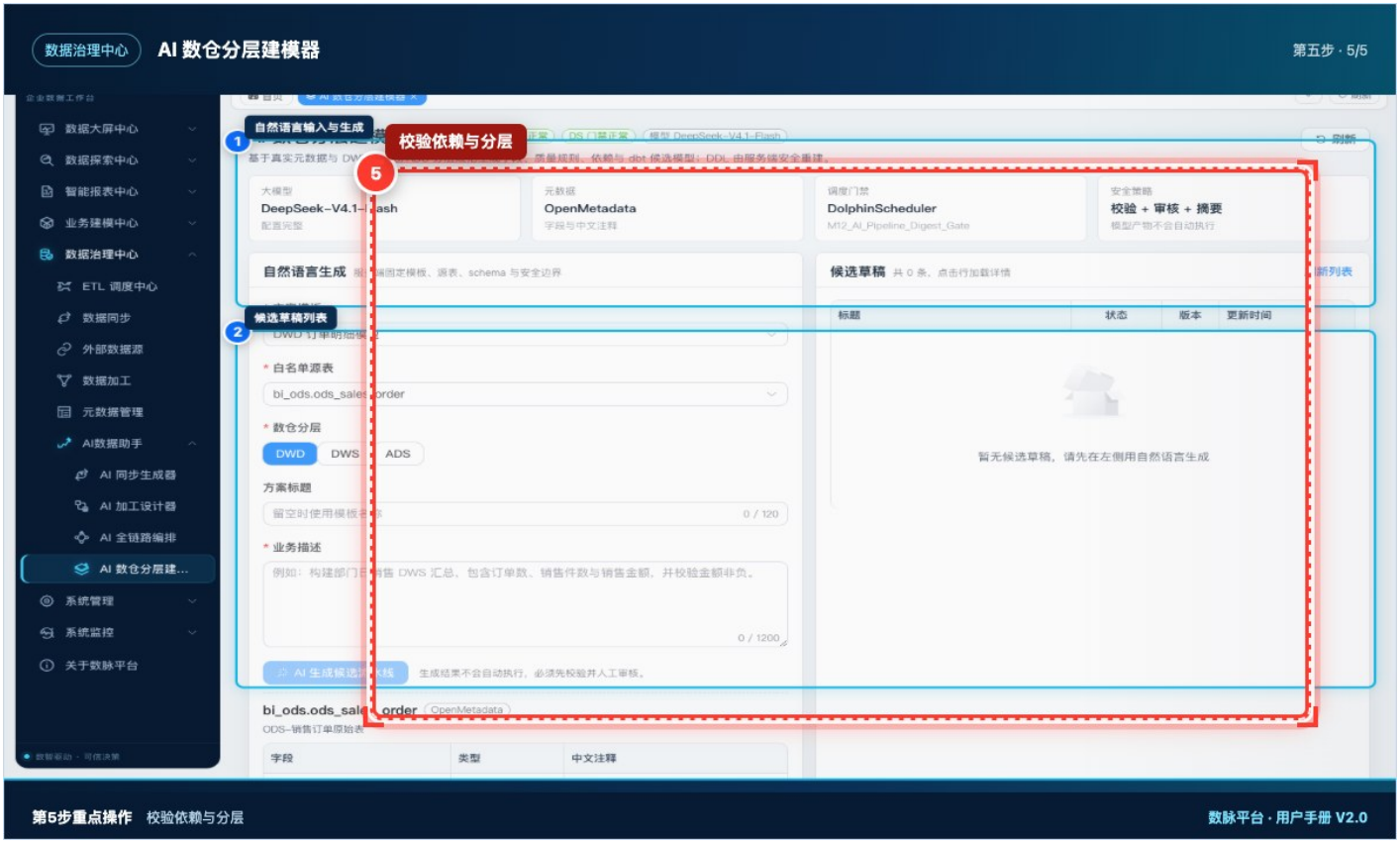
规则和结构化 DDL



步骤图 43-04 第4步：查看字段、主键、分区、度量、质量规则和结构化 DDL 。红框标出查看字段、主键、分区、度量、质量规则和结构化 DDL。

查看字段、主键、分区、度量、质量规则和结构化 DDL。

第五步：校验依赖与分层



步骤图 43-05 第5步：校验依赖与分层。红框标出校验依赖与分层。
校验依赖与分层，提交审核发布，并在发布中心确认下游数据集。

最终结果与效果
形成符合 DWD/DWS/ADS 规范的分层模型草稿，字段和质量规则可审核， DDL 由服务端重建。

操作提示： DWD 保留明细粒度， DWS 做主题汇总， ADS “
面向具体应用；选错分层会造成口径混乱。

常见问题
DDL 预览与手工 SQL 不一致。 以服务端重建结果为准，页面不允许直接提交任意 DDL。分层选错。 DWD 保留明细粒度， DWS 做主题汇总， ADS 面向具体应用；选错会导致粒度混乱。
质量规则太多导致运行慢。 优先保留主键、金额、日期和引用完整性规则，再按业务风险补充。

开发人员参考
接口： /bi/ai-pipeline/options?kind=WAREHOUSE、 /generate、 /drafts/{id}/validate|review|publish
权限： bi:etl:ai-warehouse
分层、命名、字段类型和 DDL 由服务端规范约束；模型产物不会自动执行。
审核摘要和版本记录可追溯到具体草稿、模型和发布时间。