



DATABLAU · DAM

数据血缘

产品模块用户手册

版本 9.1.2

目录

- 01 数据血缘

- 02 快速开始

- 03 血缘系统管理

- 04 应用场景

- 05 采集元数据并同步血缘资源

- 06 部署与运行检查

- 07 数据血缘探索

- 08 集成与接口参考

- 09 管理血缘空间

- 10 创建血缘任务并检查解析报告

- 11 查看和分析血缘

- 12 补充并管理手工血缘

- 13 配置命名空间、DBLink 与预处理

- 14 查询血缘图谱、加工逻辑与影响范围

- 15 在线 SQL 血缘

数据血缘

SQLink 的全称为“数据链路监测平台”，是可独立部署和授权的数据血缘产品。本栏目介绍其血缘空间、解析任务、血缘地图、应用场景、血缘编辑、影响与根因分析等专业能力。SQLink 也可作为 DAM 的血缘能力增强使用；DAM 元数据管理中的基础血缘查看仍按元数据功能使用。

产品边界

元数据管理中的基础血缘从表、字段和报表等对象出发查看上下游关系。SQLink 使用独立 SQLINK License，其专业功能不写入元数据管理章节。独立部署与 DAM 配合使用时，具体连接、同步和入口取决于实际部署配置，不能仅凭两边都有血缘视图就认定数据自动互通。

阅读前的关键概念

术语	在 SQLink 中表示什么
血缘空间	组织来源资源、血缘文件、解析任务和结果的工作范围；不是一个数据库
命名空间	区分不同系统或环境中同名对象的身份范围，帮助解析结果关联到正确的元数据
解析任务	读取指定脚本或任务文件，提取对象之间的数据加工与依赖关系，并产生解析报告
血缘地图	按图形展示对象、上下游方向和关系的探索入口；图中看到的范围受层级、深度和筛选影响
应用场景	围绕持续监测、变更影响或质量分析目标组织的一组对象和血缘关系，不复制源数据

这些概念分别对应[空间管理](#)、[任务与报告](#)、[血缘查看](#)和[应用场景](#)，不能把“已采集元数据”等同于“已生成准确血缘”。

使用者与主要入口

使用者	主要任务	常用入口
血缘消费者	搜索对象、查看上下游、分析变更影响	数据目录、血缘地图、应用场景
数据开发人员	解析 SQL、检查输入输出和加工过程	SQL 血缘、血缘地图
血缘管理员	采集元数据、解析文件、管理命名空间	系统管理、血缘管理、任务管理

SQLink 的使用顺序通常是先完成元数据采集和命名空间规划，再建设血缘链路并检查解析质量，最后开放血缘查询和应用分析。

任务入口

- [快速开始](#)：从业务目标、初始化、元数据采集到血缘应用。
- [数据血缘探索](#)：通过数据目录和血缘地图发现对象与链路。
- [查看和分析血缘](#)：搜索、链路、层级、筛选、导出和版本比较。
- [应用场景](#)：通用监测、影响分析和质量分析。
- [在线 SQL 血缘](#)：即时解析 SQL 并查看表级、字段级关系。
- [血缘系统管理](#)：采集、血缘空间、命名空间、预处理和任务管理。
- [管理血缘空间](#)：组织数据库资源、报表资源、OpenLineage 任务和解析质量。
- [采集与资源同步](#)、[血缘任务与解析](#)：建设和维护可用链路。
- [命名空间与预处理](#)、[血缘编辑](#)：处理对象身份、代码转换和人工补充关系。
- [集成与接口参考](#)：查看集成边界及手册中的接口示例。
- [部署准备与运行检查](#)：准备组件、配置服务并逐层验收。

SQLink 快速开始

本指南说明如何从一个明确的业务问题出发，完成系统初始化、元数据采集、血缘链路建设、准确性检查和应用分析。后续步骤依赖前面形成的应用系统、命名空间和元数据基础。

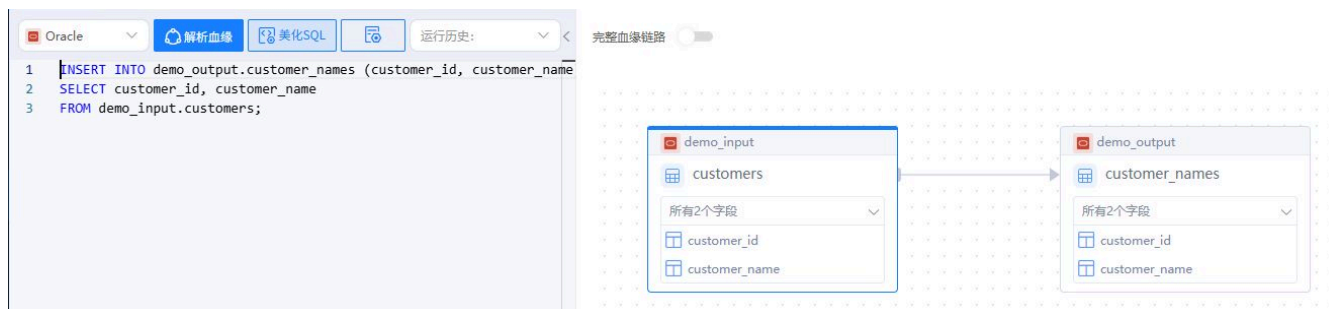
先体验一次在线解析

如果你是第一次使用 SQLink，可以先在“SQL血缘”页面完成一个小样例，认识代码与图中对象的对应关系。这个体验只解析 SQL 文本，不把 SQL 发送到数据库执行，也不替代正式元数据采集。

1. 从 SQLink 左侧导航进入“SQL血缘”。
2. 选择 Oracle 解析类型，输入下面的教学 SQL。示例中的 Schema、表和字段均为虚构名称，不代表环境中已有这些对象。
3. 初次体验不启用强制关联元数据，单击“解析血缘”。
4. 等待解析结束，检查 demo_input.customers 指向 demo_output.customer_names。
5. 展开两侧的“所有2个字段”，核对 customer_id 和 customer_name。

sql

```
INSERT INTO demo_output.customer_names
(customer_id, customer_name)
SELECT customer_id, customer_name
FROM demo_input.customers;
```



图为上述教学样例的实际解析结果；正文 SQL 换行仅为便于阅读。左侧是代码和解析控件，右侧是输入到输出的关系。图中的两张表存在于本次解析结果中，不表示已采集真实数据库元数据。

完成检查：来源是 demo_input.customers，目标是 demo_output.customer_names，箭头方向和字段名称符合语句含义。失败时先核对解析类型与 SQL 是否完整，详见[在线 SQL 血缘](#)。

只需要临时分析 SQL 的用户可继续阅读该操作页；需要长期维护真实系统链路的管理员，再按下面的建设流程执行。

快速开始流程

阶段	主要工作	完成标志
业务场景定义	设定目标、梳理系统、制定计划	明确分析问题、范围和责任
系统初始化	准备用户、机构、应用系统、数据源和标签	采集所需基础对象已建立
元数据采集	盘点资源、规划命名空间、执行采集	目标元数据进入血缘空间
血缘链路管理	导入或采集血缘文件，配置预处理并解析	形成可查询的跨系统链路
准确性检查	检查解析成功率、覆盖率和血缘质量	关键链路满足使用要求
应用分析	数据目录、血缘地图和场景分析	业务问题获得可复核结果

第一步 定义业务应用场景

设定业务目标

先描述要通过血缘解决的具体问题，例如：

- 某个表或指标变更会影响哪些任务、报表和使用者。
- 一项数据质量问题从哪个源系统或加工步骤产生。
- 一条业务链路经过哪些数据库、ETL 作业和报表。
- 哪些血缘节点无法关联到已采集的元数据。

业务目标应包含目标对象、分析方向、结果形式和验收人。没有明确目标时，不建议直接扩大采集和解析范围。

梳理关联系统

列出链路涉及的源系统、加工系统、数据仓库、报表系统和文件交换环节。对每个系统记录技术类型、环境、责任人、数据源、可用采集方式以及是否包含可解析的 SQL 或 ETL 文件。

制定行动计划

把系统按优先级分批接入。先选择一条可人工核对的代表性链路作为样本，验证命名空间、采集和解析规则后，再扩大到完整范围。

第二步 完成系统初始化

用户与机构

确认管理员、血缘消费者和数据开发人员的账号、机构及功能权限。用户能够看到菜单，不代表能够查看所有数据对象；数据范围仍以当前授权为准。

应用系统

在应用系统中登记关联系统。应用系统是血缘展示中的较粗粒度对象，也为数据源提供业务归属。系统名称和简写用于唯一识别，已被数据源或模型目录引用时不能直接删除。

数据源

建立可供采集任务复用的数据源连接。新建或修改认证信息后执行连通性测试；连接成功仅证明网络和认证可用，不代表采集范围与权限正确。

标签

按实际业务准备需要用于筛选、标识和场景管理的标签，例如临时表、核心链路或责任域。标签应有统一口径，避免同义标签影响筛选结果。

第三步 采集元数据

盘点采集对象

SQLink 可从数据源、报表系统和文件采集元数据，也可通过自定义元模型处理非标准资源。采集前确认资源类型、采集范围、采集账号权限和更新周期。

定义命名空间

命名空间为血缘节点和关系提供顶层唯一标识。同一命名空间内不应出现无法区分的同名对象。开发、测试和生产环境，以及存在重复 Schema 或对象路径的系统，应使用不同命名空间隔离。

命名空间还可以配置默认 Schema。当 SQL 中的表名未带 Schema 时，解析会使用该默认值补全对象路径。

执行采集

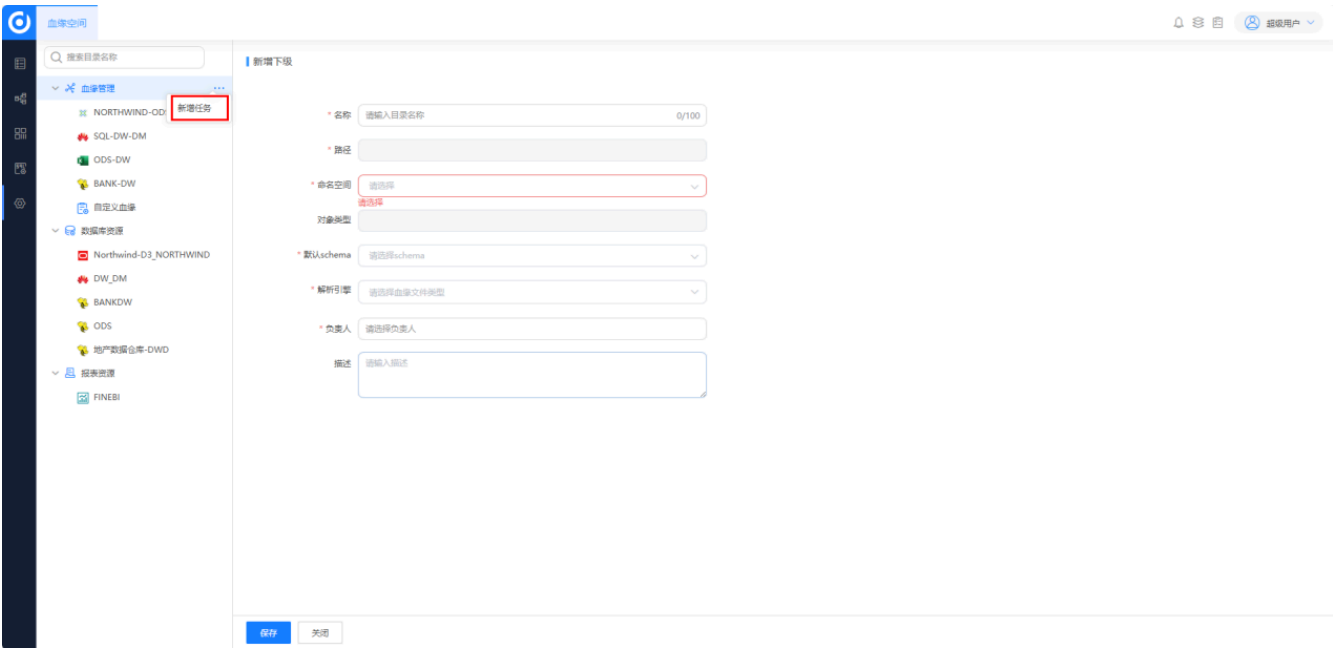
1. 创建数据源、报表或文件采集任务。
2. 选择应用系统、数据源和采集范围。
3. 为采集结果指定正确的命名空间。
4. 配置采样、名称补全或黑名单等当前任务支持的选项。
5. 保存并执行任务，观察进度和状态。
6. 在血缘空间的数据库资源或报表资源中确认同步结果。

数据源、报表、文件及资源同步的详细配置见[采集与资源同步](#)；命名空间和连接映射见[命名空间与预处理](#)。

第四步 建设血缘链路

来源	典型处理方式	关键配置
SQL 脚本	创建 Script 类型任务并导入脚本	解析引擎、命名空间、默认 Schema
ETL 文件	创建与工具匹配的 ETL 类型任务	源目标连接、命名空间映射、预处理
Excel 自定义关系	下载并填写血缘模板后导入	源目标对象路径和命名空间
数据库对象	随数据库资源同步后自动或手工解析	解析引擎、元数据限制
报表系统	随报表资源同步并解析数据准备 SQL	报表连接对应的命名空间
OpenLineage 消息	监听第三方标准事件	消息来源和任务状态

解析前检查变量、数据连接和命名空间映射。血缘文件中任何连接未关联命名空间时，任务可能解析失败。



先选择与代码来源一致的任务类型，再核对解析引擎、命名空间和导入方式。具体保存、全量/增量导入及报告检查见[血缘任务与解析](#)。

第五步 检查链路准确性

解析成功率

血缘文件可能处于成功、部分成功、解析中、失败或未解析状态。只有目标范围内的文件完成解析并处理异常后，才能据此判断链路质量。对部分成功文件查看高亮的失败代码片段和解析详情。

元数据覆盖率

覆盖率用于比较图数据库中的血缘节点与同一命名空间内数据库资源、报表资源的元数据。覆盖率不足通常表示部分血缘节点找不到对应的实际表或字段，需要检查采集范围、对象命名和命名空间。

血缘质量

使用质量分析检查孤表、环线、断点和字段一致性问题。质量检查结果是排查线索，仍需结合源代码、调度和业务流程确认根因。

第六步 开展血缘应用分析

- 在**数据目录**中按关键字、应用系统、数据源和资产类型精确找到对象。
- 在**血缘地图**中从全局视角查看不同层级和深度的数据流向。
- 在**通用场景**中持续关注核心链路并比较相似链路。
- 在**影响分析**中评估数据源、实体、指标或任务变更。
- 在**质量分析**中跟踪并处理血缘结构问题。
- 在**SQL 血缘**中即时检查 SQL 输入、输出和中间处理过程。

完成检查

- 目标系统、数据源和命名空间与实施计划一致。
- 元数据已同步到血缘空间，关键对象可以搜索到。
- 血缘文件的失败和部分成功记录已处理或登记。
- 样本链路的起点、终点、中间对象和方向经责任人核对。
- 用户能够按授权查看链路和报告。
- 场景分析结果能够回答最初定义的业务问题。

相关文档

- [数据血缘探索](#)
- [查看和分析血缘](#)
- [应用场景](#)
- [血缘系统管理](#)

SQLink 血缘系统管理

本指南组织元数据采集、血缘空间、解析任务、预处理、血缘编辑和运行调度。首次建设先按实施顺序准备，日常操作进入相应详细指南。

任务	详细指南
采集数据库、报表、DDL、API 或自定义对象	采集与资源同步
创建解析任务、导入文件并检查报告	血缘任务与解析
设置对象身份、连接映射及预处理	命名空间与预处理
补充有明确依据的手工关系	血缘编辑
管理执行周期、节点与日志	任务调度

建议实施顺序

1. 梳理系统、环境、数据源和代码来源。
2. 规划命名空间及默认 Schema。
3. 配置元数据采集并检查采集结果。
4. 将数据库、报表或 OpenLineage 资源纳入血缘空间。
5. 根据代码来源创建解析任务，必要时配置预处理和连接映射。
6. 执行任务并处理解析报告中的失败或缺失项。
7. 对自动解析无法覆盖且有证据的关系进行受控编辑。
8. 配置调度并持续检查成功率、覆盖率和链路质量。

采集管理

数据源采集

数据源采集为血缘关联提供数据库元数据。创建采集配置时选择数据源和采集范围，并根据实际需要设置采样、名称补全和黑名单。黑名单用于排除无需采集的对象，使用前应确认不会排除关键上下游。

执行后在任务列表查看进度和状态，再抽样检查数据库、Schema、表和字段是否完整。采集成功不代表所有业务对象均已覆盖。

报表采集

报表采集用于获取报表资源及其数据集等信息。保存前执行连接测试；正式采集通常由后台任务完成，应在任务记录中检查完成状态、资源数量和失败信息。

文件采集

文件采集可处理 DDL 等文件来源。配置时选择解析引擎、命名空间和默认 Schema；若文件通过 API 或外部流程提交，还需核对来源标识和元数据上下文。不同数据库语法应使用匹配的解析引擎。

自定义元模型采集

当标准采集无法表达特定对象时，可依据已配置的自定义元模型采集。采集前确认元模型的类型、属性和关系定义，采集后检查对象是否落入预期类型，避免形成无法参与分析的孤立节点。

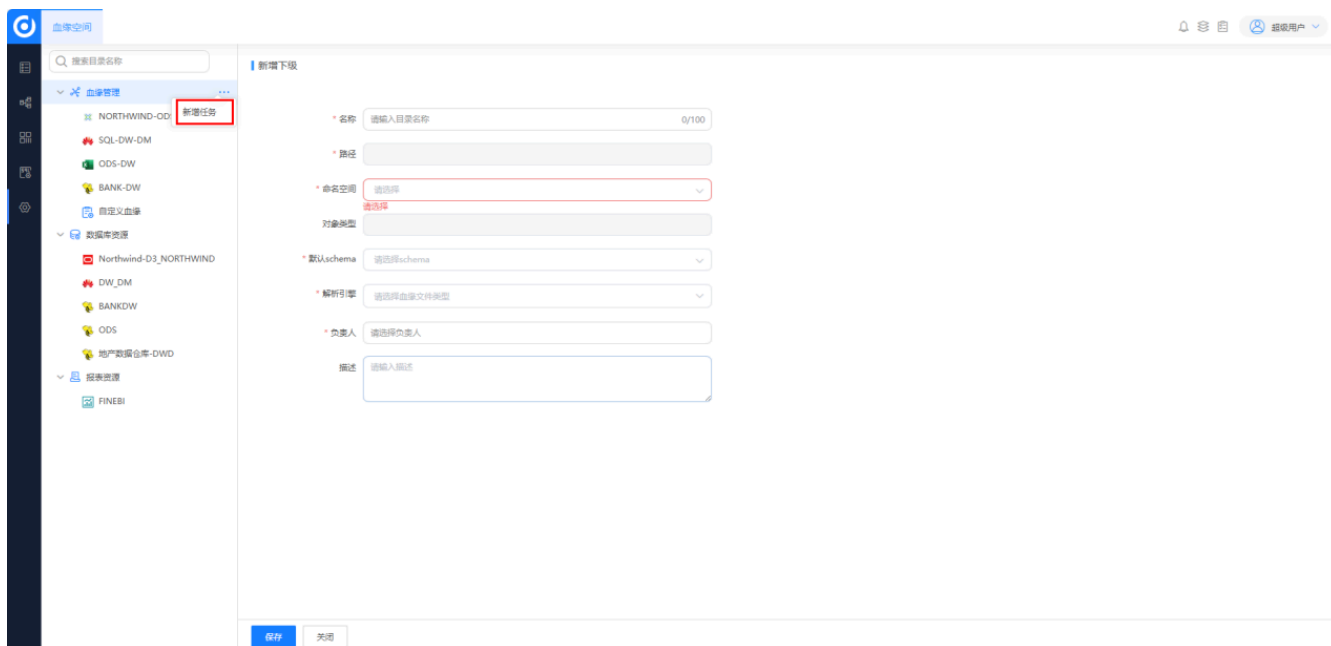
血缘空间

血缘空间集中管理参与血缘分析的资源，主要包括以下部分：

区域	管理内容	核对重点
血缘管理	已解析关系及其状态	对象身份、方向、层级和解析结果
数据库资源	已采集的数据库元数据	命名空间、Schema 和覆盖范围
报表资源	已采集的报表及数据集	报表来源和数据库对象关联
OpenLineage 任务	外部运行任务与事件	作业标识、输入输出和事件完整性

元数据采集完成后先在血缘空间检查资源，再创建正式血缘任务。资源缺失时应回到采集配置处理，而不是通过手工血缘长期弥补。

创建血缘任务



血缘任务按来源可分为：

- **Script**：处理脚本或 SQL 文件。
- **ETL**：处理受支持的 ETL 工具导出内容或任务定义。
- **Container of User-Defined-Object**：血缘编辑保存后生成的对象容器，不可再次解析。

创建时应记录任务用途、责任人、来源、解析引擎、命名空间和运行方式。先用小批量文件验证，再导入完整范围。

全量与增量导入

全量导入将指定目录中的脚本导入，重复文件被覆盖并变为“未解析”，支持多层目录和不同层同名文件。增量导入按内容、大小、路径和文件名比较，相同文件跳过，不同文件更新。执行前核对范围，不把“全量”理解为已有文件必然全部删除。

强制绑定元数据

强制绑定用于把代码解析得到的对象关联到已采集元数据。启用前必须确认命名空间、数据连接、默认 Schema 和对象命名规则一致。绑定失败时先修复上下文，不应按名称随意选择对象。

预处理管理

预处理在正式解析前处理变量、连接信息或不规则文本。可用方式包括：

- **JavaScript**：按受控脚本转换输入内容。
- **正则表达式**：匹配并替换格式稳定的文本或变量。
- **大模型**：辅助处理难以用固定规则覆盖的内容，结果必须抽样复核。

规则可引用任务变量和数据连接。配置后先比较预处理前后的文本，确认没有改变业务语义，再提交解析。预处理脚本或表达式应由授权人员维护并保留变更记录。

解析报告

任务执行后，解析报告通常按以下结果分类：

结果	含义	处理方式
成功	内容完成解析	抽样核对关键链路
部分成功	仅部分语句或对象完成解析	查看未解析明细并修复后重跑
语法错误	内容不符合所选解析语法	核对引擎、代码完整性和变量
元数据缺失	解析节点无法关联到元数据	补采元数据或修正命名空间与 Schema
其他错误	运行环境或未知原因导致失败	查看日志并记录复现样本

解析成功率衡量输入内容的处理情况，元数据覆盖率衡量解析对象与已采集元数据的关联情况。两个指标应分别检查。

DBLink 映射

当代码通过 DBLink 等机制引用其他数据源时，配置映射把代码中的连接标识对应到正确的数据连接和命名空间。映射错误会造成跨系统链路断开或关联到错误环境。

命名空间管理

命名空间用于唯一识别不同系统和环境中的对象。规划时记录所属系统、环境、资源类型和命名规则。血缘空间允许更换任务命名空间，但会影响对象匹配，执行前检查相关采集与血缘关系。

血缘编辑

血缘编辑用于补充有明确证据但无法自动采集的关系。编辑中的删除仅作用于本次新增实体和连线，不删除原画布内容。保存为自定义血缘文件后可查看、删除，但不能再次编辑修改；具体步骤见[血缘编辑](#)。

手工编辑应记录来源证据和原因。若同类关系持续需要人工维护，应优先修复采集、预处理或解析配置。

任务管理与调度

任务管理用于查看运行状态、执行历史和计划。配置调度时选择时间模板；多个模板的可执行日期取交集，再按重复周期执行。被引用模板删除或禁用后，任务不再受该模板约束。

日常监控至少关注最近和下次执行时间、成功或失败状态、处理文件数量、解析报告和连续失败次数。修改调度后执行一次受控验证，确认任务真正触发。

数据源、驱动与插件

新增数据源前确认网络、账号权限、数据库类型和驱动兼容性。驱动或解析插件的增加、替换和停用属于管理员操作，应先在测试范围验证，并记录受影响的数据源和任务。不要在用户手册或截图中保存连接密码。

结果验证

- 数据库、报表和文件资源均进入正确命名空间。
- 采集范围与计划一致，黑名单未排除关键对象。
- 解析报告中的部分成功和失败项已有处理结论。
- 成功率与元数据覆盖率分别达到项目验收要求。
- 跨数据源链路的连接和 DBLink 映射正确。
- 手工血缘有证据、操作者和时间记录。
- 调度已实际触发，执行历史与预期时间一致。

故障排查

现象	优先检查
元数据未进入血缘空间	连接、采集范围、黑名单、任务状态和数据权限
文件解析失败	文件完整性、解析引擎、变量、预处理和错误日志
出现大量元数据缺失	命名空间、默认 Schema、采集范围和对象大小写
跨系统链路断开	数据连接映射、DBLink 映射和上下游任务是否均已采集
出现重复节点	命名空间规划、对象标识和重复导入方式
调度未执行	启停状态、时间模板交集、依赖任务和调度日志

相关文档

- [SQLink 快速开始](#)
- [数据血缘探索](#)
- [查看和分析血缘](#)
- [应用场景](#)
- [在线 SQL 血缘](#)

SQLink 应用场景

“应用场景”是 SQLink 为特定分析目标保存的对象范围和查看方式，把数据目录中的数据源、数据实体和指标组织到一个可持续分析的业务范围中。它不是数据开发任务，也不复制数据或替代采集和解析；场景使用已经进入 SQLink 的元数据与血缘关系回答特定问题。

使用前提

- 目标数据源、实体或指标已进入数据目录。
- 关键血缘已完成解析，并经过抽样检查。
- 当前用户对场景对象具有查看权限；创建、编辑和删除权限由管理员分配。
- 已明确场景负责人、分析目标和处理结果的接收人。

场景中的对象

创建场景时可从数据目录选择数据源、数据实体和指标。对象是否可选取取决于采集范围、数据权限和当前目录内容。添加前应核对对象所属系统和命名空间，避免选择同名但不同环境的对象。

场景支持关注和分享。关注用于持续查看本人关心的场景；分享前应确认接收人具有场景内对象的数据权限，分享操作不会扩大其原有权限。

通用场景

通用场景适合围绕一条核心业务链路持续观察相关对象和变更事件。“核心链路”指本场景重点关注的对象及其上下游依赖，不表示 SQLink 会自动判断哪条链路在业务上最重要。

创建场景

1. 进入应用场景，选择新建通用场景。
2. 填写能够说明业务范围的名称和描述。
3. 从数据目录添加需要关注的数据源、实体或指标。
4. 核对对象所属系统、环境和上下游范围。
5. 保存场景，进入场景详情检查核心链路。

创建后，具有相应权限的用户可继续编辑范围、删除场景或进入管理视图。删除前应确认场景不再承担持续监测或协作任务。

核心链路与变更事件

在场景详情中检查核心对象及其上下游关系。发生对象或链路变更时，可在场景中查看相关事件并判断是否需要通知责任人。场景内出现的事件是排查入口，仍需结合任务日志和实际发布记录确认变更原因。

影响分析

影响分析用于变更前评估和变更后核对。分析对象不同，检查重点也不同。

分析对象	主要用途	重点检查
数据源	评估数据源结构或采集范围变化	受影响实体、任务、报表和下游系统
数据实体	评估表、视图等对象变化	字段级或实体级下游影响
数据指标	评估指标口径或依赖变化	指标来源、计算链路和消费对象
任务	评估调度或转换任务变化	输入、输出和依赖任务

事前影响分析

1. 选择计划变更的数据源、实体、指标或任务。
2. 设置分析方向、层级和深度。
3. 检查受影响的直接下游和完整下游链路。
4. 识别关键系统、报表、任务和责任人。
5. 保存或导出分析结果，交由相关负责人复核。
6. 在变更方案中记录风险和回退安排。

事后影响核对

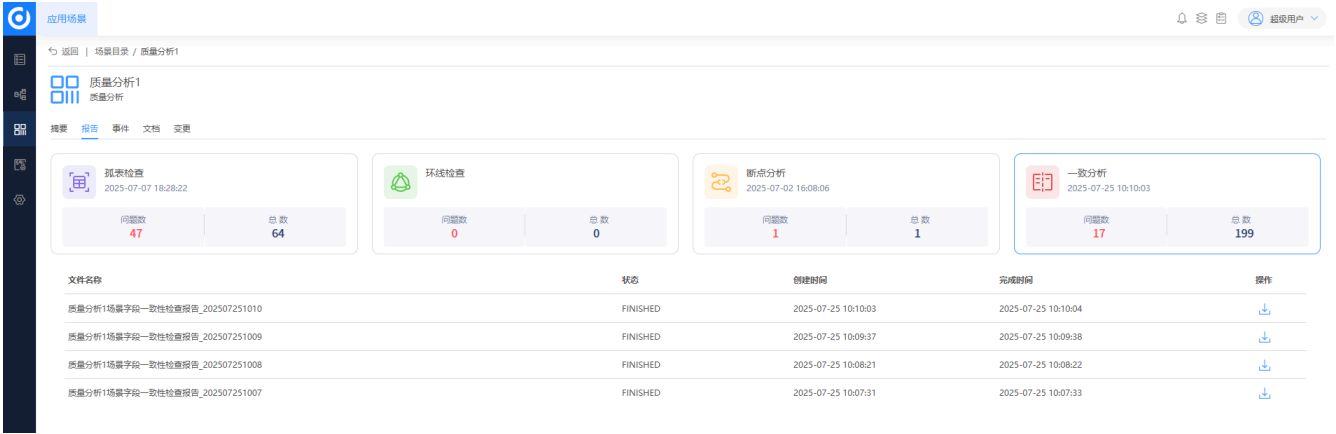
变更完成后重新运行相同范围的影响分析，对比新增、消失或方向变化的关系。若结果与事前评估不一致，先检查采集和解析是否已更新，再判断是否为实际变更遗漏。

影响广播

“影响广播”是把已识别的潜在受影响对象或范围通知相关人员的协作动作，不是血缘计算本身。系统可基于已识别的影响范围辅助发起广播。自动广播与手工广播的触发条件、接收范围和发送方式可能不同，使用前应由管理员确认当前配置。广播只用于告知和留痕，不能替代正式变更审批或责任人确认。

质量分析

质量分析从血缘结构检查孤表、环线、断点以及上下游对象的一致性。



孤表分析

孤表是未形成预期上下游关系的实体。发现孤表时依次检查：对象是否应参与数据加工、相关代码或任务是否已采集、命名空间是否一致，以及解析是否成功。业务上确实独立的表不应被误判为缺陷。

环线分析

环线表示链路中存在循环关系。循环可能来自实际迭代处理，也可能由对象映射或解析错误造成。应结合任务顺序、代码和调度记录判断，不要仅凭图形删除关系。

断点分析

断点表示预期连续的链路在某处中断。重点检查上游输出与下游输入的名称映射、临时表、跨数据源连接、脚本变量和缺失的采集任务。

一致性分析

一致性检查关注上下游字段的名称、数据类型和长度。发现差异时需判断这是合理转换还是潜在风险，并结合转换代码确认。类型或长度不同不必然代表错误。

结果验证

- 场景名称和描述能清楚说明业务问题。
- 场景对象来自正确的系统、环境和命名空间。
- 影响分析的方向、层级和深度符合变更范围。
- 广播接收人和通知边界已经人工复核。
- 孤表、环线、断点和一致性问题均有调查结论或处理人。
- 修复采集、映射或代码后已重新运行分析并比较结果。

常见问题

找不到要添加的对象

检查对象是否已完成元数据采集、是否进入数据目录，以及当前用户是否具有查看权限。不要用名称相近的其他环境对象代替。

影响范围明显过小

检查分析深度和筛选条件，再检查相关解析任务、命名空间及元数据覆盖率。未发现影响不等于不存在尚未采集的链路。

质量问题数量突然增加

先核对最近的采集、解析或映射变更，再抽样检查问题对象。不要在确认原因前批量手工补线。

相关文档

- [查看和分析血缘](#)
- [数据血缘探索](#)
- [血缘系统管理](#)

采集元数据并同步血缘资源

SQLink 先采集对象结构，再把资源同步到血缘空间供解析匹配。数据源连接可以复用于不同 Schema 的采集，保存连接不等于采集完成。

数据源采集

1. 在 SQLink“系统管理 → 数据源管理”准备应用系统与数据源，测试连接。
2. 进入“血缘管理 → 采集管理 → 新建采集 → 数据源”。
3. 填写名称、数据源及采集范围。
4. 配置采样、元数据名称补全和黑名单，确认不会排除关键对象。
5. 确定保存后，在任务列表选择“执行”。
6. 查看采集进度与结果，到数据目录及血缘空间核对对象。

报表采集

新建采集选择“报表”，选对应工具类型，填写服务 URI、认证及报表目录等表单要求，测试连接后确定。采集在后台执行，需检查任务状态与报表、数据集和数据库对象关联。

报表资源中的数据库连接需要映射到正确命名空间后再解析，不能仅依靠报表名称关联库表。

DDL 文件与 API

新建采集选择“文件”。DDL 脚本的数据库选择 DDL Scripts，设置命名空间、解析引擎和默认 Schema。解析引擎用于判断脚本语法，命名空间用于存储和匹配对象；两者数据库类型允许不同，但应与实际语义一致。

API 元数据使用内置 REST API 元模型，可通过 JSON 文件或直接 HTTP 访问方式采集。选择方式后核对对象及关联，不将普通文本接口说明当作结构化定义。

自定义元模型

新建文件采集，数据库类型选择“元模型”，选择已建立的模型与命名空间，下载模板，填写对象后上传并运行。完整结构及填表差异见[元模型设计](#)和[元模型导入](#)。

数据库与报表资源

采集后自动在“血缘空间”的数据库资源或报表资源中建立任务。数据库资源包括表字段及视图、存储过程等可解析内容；报表资源包括报表结构及数据准备 SQL。

资源中的元数据不能直接编辑或删除，应通过采集同步更新。可调整解析引擎并重新解析。

重新同步元数据会先清除任务原有元数据及血缘信息。 仅在确定需要同步且来源可用时执行，完成后检查资源与链路恢复情况。更换命名空间前同样需要确认关联影响。

OpenLineage

“OpenLineage 任务”接收第三方通过 Kafka 提供的运行事件，管理运行状态和作业父子关系。字段血缘事件可保存为关系，事件中的 SQL 可提取为血缘文件。

实施时先核对实际 Kafka 主题、事件格式、生产者配置和命名空间，再检查输入输出、运行状态及生成关系。实时事件接入不等同于数据库采集；缺少源目标事件时不能推定完整链路。

SQLink 部署准备与运行检查

本页供实施和运维人员准备部署、核对配置并验收服务。具体安装包、组件组合、资源大小及启动脚本使用项目交付清单；部署成功后再进行[元数据采集和链路建设](#)。

准备部署清单

部署前将以下内容落实到具体机器和负责人。

项目	应准备的内容	检查方法
部署形态	独立部署或与 DAM 集成，实际服务及节点分布	对照交付架构逐项登记，不混用两套拓扑
主机与存储	主机名、地址、部署用户、安装目录、数据及日志目录	部署用户可读写所需目录，容量满足项目计划
数据库	数据库实例、字符与大小写配置、应用账号、备份位置	使用应用账号验证连接及初始化所需权限
服务配置	Nacos 地址、命名空间、各应用配置和注册地址	核对客户端实际可达的地址，不填容器内部不可达地址
网络	浏览器入口、服务间通信、采集源连接	按实际配置逐段检查，不仅检查首页端口
许可	项目要求的许可文件及关联机器信息	由管理员核对许可内容与部署机器
安装材料	组件包、服务包、前端资源、配置模板、启动脚本	归档交付清单，避免混用不匹配的包

部署用户与运行目录应保持一致。不要因为参考脚本使用某目录，就在现有生产机器上覆盖同名目录。

组件与服务职责

组件或服务	职责	重点检查
关系数据库	保存后台业务数据	实例、账号、库名及大小写行为
Nacos	服务注册与配置管理	命名空间、配置内容和服务注册状态
Redis	缓存	连接地址、认证及运行用户
Elasticsearch	全文搜索	索引服务状态、运行权限及资源限制
Nebula	图数据存储	图服务、存储节点注册及节点可用状态
Kafka	消息处理	服务连接与配置一致性
用户、基础服务	用户身份与公共能力	数据库和配置中心连接
元数据、调度、SQLink 服务	元数据、任务执行和血缘功能	服务启动日志、注册与相互调用
应用网关、Nginx	后端路由及前端入口	转发目标、静态资源目录及访问路径

这里的“应用网关”承担平台服务路由，与 DDS 中执行动态脱敏的“安全网关”不同。数据标准等共享服务是否部署，按所采用的集成方案处理。

按顺序安装与配置

1. 准备数据库实例及部署账号，核对应用初始化所需权限。
2. 先安装交付清单中的基础组件，逐个确认启动成功。Nebula 还需检查存储节点已注册且可用。
3. 将各微服务包展开至约定目录，核对应用包与启动脚本匹配。
4. 配置启动脚本中的本机注册地址、服务端口和 Nacos 地址，按实际资源设置 JVM 内存。
5. 在 Nacos 对应命名空间编辑并发布配置，逐项核对数据库、组件及服务连接。
6. 以约定运行用户启动用户管理、基础服务及所需业务服务。每启动一项，就检查进程、监听、日志及注册状态。
7. 部署前端静态资源，配置 Nginx 的静态目录与服务转发；核对 SQLink 入口与 API 路由。
8. 进入产品，完成许可、角色、数据源等初始化，再执行下述验收。

配置中心中常见应用配置包括 datablau_usermgmt.yaml、datablau_base.yaml、datablau_metadata.yaml、datablau_sqlink.yaml 和 datablau_gateway.yaml。使用安装包实际提供的配置名称和命名空间，不凭名称相似另建一份导致服务读不到配置。

分层验收

层次	验收动作	成功判据
进程	检查各组件和微服务	持续运行，没有启动后退出或反复重启
注册	检查 Nacos 服务状态	注册实例地址、端口正确，应用可互相访问
入口	登录并打开 SQLink	页面和静态资源正常，业务请求可返回
元数据	用受控数据源执行小范围采集	任务成功，能找到预期 Schema、表和字段
血缘	创建测试任务并解析已知脚本	报告可查看，关键上下游与脚本一致
调度	配置一个受控计划并检查历史	在预定时间触发并保留结果
权限	使用实际业务角色查看对象与入口	可见范围和操作符合授权

首页可打开只证明入口可用；不能替代采集、解析与调度验收。验收时保留环境配置摘要、任务标识、时间、输入样本和检查结论。

备份与恢复准备

部署资料要求定期备份数据库。实施时还应清点当前方案中的图数据、上传文件、服务配置及前端资源，将备份范围、周期、保留时间和恢复顺序纳入运维方案。不能把单个数据库备份成功理解为整个系统已可恢复。

恢复演练先在隔离环境执行，再检查登录、对象数量、关键链路和任务配置。恢复后先核对调度状态，避免原定任务立即访问不正确的数据源。

常见启动问题

现象	排查顺序
Elasticsearch 提示不能使用 root 启动	改用交付方案约定的普通运行用户，检查目录归属
配置文件 AccessDeniedException	核对运行用户对配置与数据目录的权限，不直接放开所有目录权限
文件描述符或虚拟内存映射限制不足	由运维按错误消息及组件要求调整系统限制，重新验证
微服务未注册或调用失败	检查 Nacos 地址、命名空间、注册地址、端口与网络
前端正常但接口报错	检查 Nginx 转发、应用网关和目标微服务日志
采集正常但血缘无法展示	检查 SQLink 与图服务、任务解析结果及命名空间关联

相关操作：[采集与资源同步](#)、[血缘任务与解析](#)、[运行调度](#)。

数据血缘探索

SQLink 提供数据目录和血缘地图两条探索路径。数据目录从已经采集的元数据出发，适合精确寻找一个表、字段、报表或指标；血缘地图从已经解析的血缘对象出发，适合从全局视角浏览数据流向。

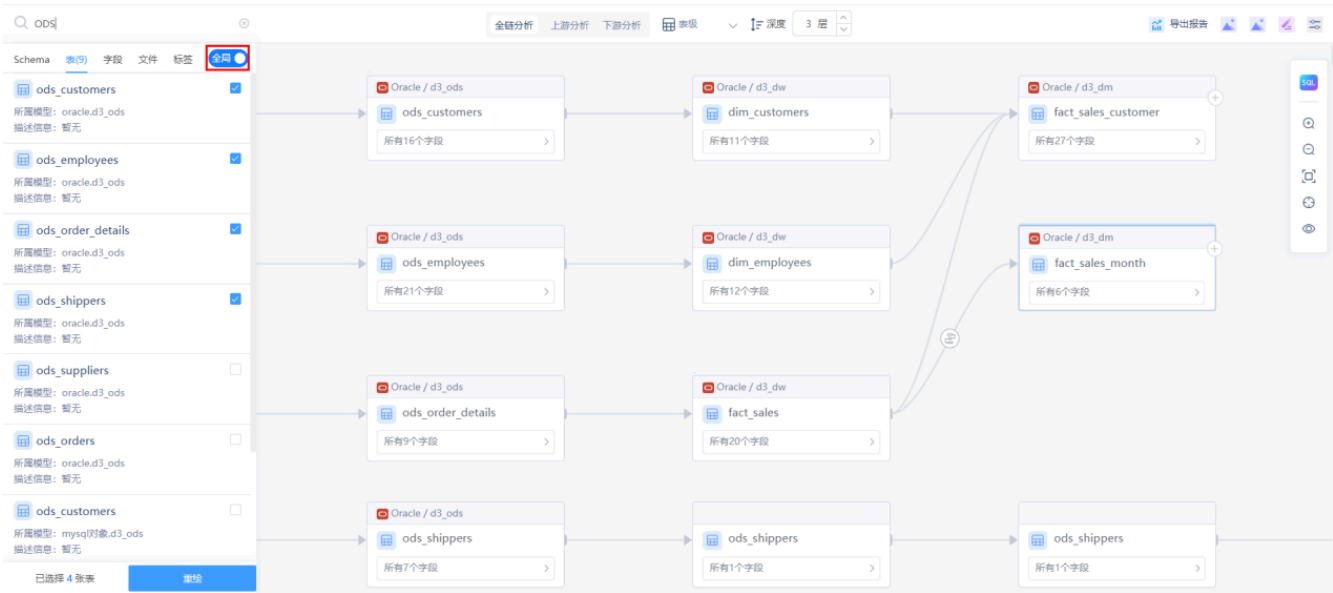
数据目录与血缘地图的区别

对比项	数据目录	血缘地图
主要数据	采集到的元数据	解析后保存的血缘对象
典型用途	精确查找已知对象	浏览整体链路和跨层级关系
常用筛选	关键字、应用系统、数据源、资产类型、标签	全局搜索、图内搜索、层级、深度、标签
展示特点	包含较丰富的中文名、字段类型等元数据信息	简化对象信息，突出流向和关系
可继续操作	查看血缘、加入应用场景	钻取、筛选、影响分析和报告导出

使用数据目录

入口：导航栏 **数据目录**。

1. 输入对象名称或关键字进行模糊搜索。
2. 按应用系统、数据源和资产类型缩小范围。
3. 需要时增加标签或数据所有者条件。
4. 打开目标对象，核对对象类型、路径和所属系统。
5. 查看该对象的血缘，或将对象加入应用场景。



识别元数据与血缘不匹配

从数据目录进入血缘时，系统会区分元数据和血缘对象是否能够对应。

情况	页面表现	处理建议
表元数据存在，但没有对应表血缘	仍显示元数据及字段，链路可能缺失	检查血缘文件是否已采集和解析
表血缘存在，但没有对应表元数据	表以差异样式显示，字段类型可能使用默认图标	检查元数据采集范围和命名空间
字段元数据存在，但没有字段血缘	字段可在元数据抽屉中查看，但无对应关系	检查解析粒度和 SQL 内容
字段血缘存在，但没有字段元数据	字段以差异样式显示	检查字段采集、对象命名和映射

这种差异通常反映采集范围、解析内容或对象映射之间尚未对齐。

使用血缘地图

入口：导航栏 **血缘地图**。

1. 使用全局搜索选择一个或多个 Schema、实体或属性。
2. 选择上游、下游或全链路分析。
3. 从 Schema 级逐步钻取到实体级或属性级。
4. 调整血缘深度，控制加载的关系范围。
5. 使用图内搜索定位当前画布中的对象。
6. 使用标签、实体类型等筛选条件隐藏非关注对象。

血缘地图中的对象来自解析结果，可能与数据目录中的元数据对象数量不同。分析前先确认当前对象、命名空间和所属系统。

选择正确的探索路径

- 已知具体表、字段、报表或指标名称时，从数据目录开始。
- 需要了解平台总体数据流向时，从血缘地图开始。
- 需要组织固定范围并持续观察时，将数据目录对象加入应用场景。
- 发现对象差异或断链时，转由管理员检查血缘空间、命名空间和解析详情。

结果验证

- 搜索结果对应正确系统和环境，不是同名对象。
- 数据目录显示的元数据属性与采集结果一致。
- 血缘地图的链路方向与实际加工过程一致。
- 对象不匹配情况已记录并分派采集或解析责任人。

相关文档

- [查看和分析血缘](#)
- [应用场景](#)
- [血缘系统管理](#)

SQLink 集成与接口参考

本页面向需要把血缘能力接入其他应用的开发者。接口路径与参数依据产品手册整理，调用前应与目标服务实际接口定义核对。示例不包含真实地址、凭据或对象 ID。

DAM 内集成

DAM 的血缘空间、预处理和命名空间管理可使用 SQLink 的相应管理能力；从元数据对象进入的血缘与独立“数据血缘”栏目分属不同授权范围。后者包含数据目录、血缘地图、应用场景及在线血缘，参见[许可边界](#)。

元数据对象的血缘页面可通过嵌入方式呈现，但嵌入本身不会授予额外权限。

请求约定

基础路径为部署服务下的 /sqlink，下表路径追加在此之后。所列接口使用 POST。默认请求、响应为 JSON；文件上传使用 multipart/form-data，模板下载返回二进制内容。

认证继承系统统一认证。实际认证头、Token 获取和有效期应按当前部署配置，不在此虚构固定认证头。失败时区分参数错误、认证失败、路径或对象不存在、服务异常；HTTP 成功还需核对业务结果。

命名空间、标签与自定义关系

操作	路径	主要请求内容
创建命名空间	/dataArea/createDataArea	Body: 必填 name、objectType；可选 defaultSchemas
查看元数据覆盖率	/dataArea/getMetadataCoverage	Query: 必填 namespaceId (Long)
批量删除命名空间	/dataArea/deleteDataArea	Body: 命名空间 ID 数组；有相关血缘文件时不能直接删除
创建自定义血缘	/customLineage/create	Body: name、edges 必填，description 可选；每条边包含 startId、endId
为对象添加标签	/tag/addTag	Body: objectId、tagIds；可配置扩散
查标签关联对象	/tag/getByTag	Query: 必填 tagId (Long)

标签扩散参数包括 diffusion、direction、depth。方向取 FORWARD（下游）、BACKWARD（上游）或 ALL；扩散默认关闭，启用后应明确深度与实际影响范围。

查询图谱、加工逻辑和影响

操作	路径	请求与结果
查询图谱	/lineageGraph/query	Body 指定节点和查询条件，返回子图
获取加工逻辑	/lineageGraph/getTransform	Body 使用图查询条件并指定 id 或 queryIds；Query 可用 isGraph 控制图内模式
获取影响结果	/lineageGraph/getImpact	Body 指定节点、方向、深度及资源类型

参数类型、必填项、默认值、响应字段以及各接口的方向枚举差异，见[查询接口详解](#)。解释结果前核对方向、粒度及是否排除了临时对象；不能把图查询的全部参数直接套用到影响分析。

文件、SQL 与断链任务

操作	路径	请求内容
上传文件	/lineage/uploadLineageFile	Query: lineageEngine、objectType、folderId 必填，defaultSchema、parse 可选；Body: 二进制 file
解析 SQL	/lineage/parse	Body: 必填 sqlContent，可选 defaultSchema、sqlType
发起断链分析	/lineage/executeBrokenLineageJob	Body 可限定 sourceNamespaces、targetNamespaces、direction

SQL 请求结构示例：

json

```
{
  "sqlContent": "INSERT INTO demo_target.orders SELECT order_id FROM demo_source.orders",
  "defaultSchema": "demo_target",
  "sqlType": "DML"
}
```

demo_source 和 demo_target 是示例 Schema，需替换为获准使用的对象上下文。解析请求与实际执行数据库 SQL 是不同操作。上传返回文件 ID后还需检查是否解析、解析报告及对象匹配。

预处理脚本与调度

操作	路径	请求内容
创建脚本	/script/createScript	Body: name、scriptType、objectType、scriptContent; 可选 enabled
下载脚本模板	/script/exportScriptTemplates	返回可用于导入的 Excel 模板
创建任务	/job/createJob	Body: 名称、绑定资源、调度与执行类等任务配置
立即运行	/job/runJob	Query: jobId; 可选 Body 任务描述符覆盖默认参数

scriptType 使用 JAVASCRIPT、REGEX 或 LLM。创建任务需要 name、bindResourceId、bindResourceType、schedule、className，可带 description 和 jobContent。绑定类型和执行类应从适配任务定义取得，不自行猜测 Java 类名。

立即运行返回任务实例标识，不等于任务完成，应通过任务执行记录检查结果。删除命名空间、标签扩散、上传自动解析和任务执行均会改变服务数据，集成测试应明确范围并避免重复提交已成功的操作。

管理血缘空间

血缘空间是 SQLink 组织血缘资源、解析任务、命名空间和版本结果的工作区。它不是单个数据库，也不是元数据目录；一个空间可以承载多个来源，但必须用清晰的目录和命名空间隔离范围。

空间规定“哪些资源和任务放在一起管理”，命名空间规定“解析到的同名对象归属于哪里”。两者用途不同：把文件加入空间后，仍需检查对象名称是否映射到正确命名空间和元数据。

血缘空间的四类资源

资源	作用	典型来源
血缘管理	保存 SQL、ETL、Excel 等待解析文件和解析结果	SQL Script、ETL 文件、自定义关系
数据库资源	保存采集到的数据库对象和元数据模型	数据库采集、DDL 文件
报表资源	保存报表、数据集及报表准备 SQL	报表采集
OpenLineage 任务	接收运行事件和作业父子关系	Kafka/OpenLineage 事件

这些资源的任务可以手工创建，也可以由元数据采集任务自动同步。数据库资源和报表资源中的采集对象通常不直接编辑，来源变化应回到采集任务处理。

表中的“血缘管理”是 SQLink 空间内保存和解析血缘文件的资源分区，不是 DAM 元数据模块的“血缘管理”License；二者不能据名称相同而视为同一授权或操作入口。

规划空间和目录

创建空间前确定业务范围、来源系统、环境、责任人和解析引擎。建议按业务域或环境建立空间，再在空间内按来源和用途建立目录；不要把开发、测试和生产脚本混在一个无边界目录。

空间命名应包含业务范围和用途，目录名称应能对应责任团队。需要多人维护时，先按系统管理配置角色、空间权限和任务权限，再开放写操作。

新建血缘任务

1. 进入“血缘空间 → 血缘管理”，选择“新建任务”。
2. 选择任务目录、解析引擎和导入方式。
3. 配置全量或增量导入策略，指定文件目录和元数据限制。
4. 绑定命名空间、默认 Schema、DBLink 和必要的预处理。

5. 保存任务后先执行小批量测试，确认报告再扩大范围。

全量导入会处理目录下所有文件；增量导入按文件内容、大小、路径和名称比较，仅处理发生变化的文件。“强制绑定元数据”要求解析出的表或字段能够关联到已采集的元数据，而不是只生成未绑定的图节点；启用后任一对象找不到对应元数据都可能导致任务失败，应在正式运行前检查覆盖率。

数据库资源

数据库资源通常由数据库采集或 DDL 采集任务创建。运行后检查数据库、Schema、表和字段数量，确认对象与命名空间一致。采集成功不等于血缘解析成功，必须继续执行对应解析任务并查看报告。

报表资源

报表资源由报表采集任务同步生成，包含报表、数据集和准备 SQL。报表资源的元数据不应直接编辑或删除；如解析失败，检查报表数据连接对应的命名空间和解析引擎，再重新解析。

OpenLineage 任务

OpenLineage 任务接收第三方运行事件，管理 RunEvent 时间、状态、Job 父子关系和字段血缘。先确认消息来源、Job 标识和事件顺序，再检查事件是否生成图关系。事件缺失时，不要用手工关系补写成真实运行链路。

命名空间与 Schema

命名空间把文件中的对象名称映射到实际元数据。解析前应配置数据库类型、默认 Schema、连接和 DBLink 映射；同一个表名在不同系统中重复出现时，必须依赖命名空间区分。

常见错误包括：对象未绑定命名空间、默认 Schema 错误、DBLink 未映射和连接名称与采集名称不一致。错误时先在对象详情中确认命名空间和 Schema，再回到[命名空间与预处理](#)修正。

预处理和解析报告

预处理用于在解析前标准化变量、修正脚本格式或提取嵌入 SQL。可用方式取决于部署配置，常见类型包括 JavaScript、正则表达式和 AI 模型。预处理会改变解析输入，正式环境应保留原文件并记录脚本版本。

解析报告至少区分成功、部分成功、语法错误、元数据缺失和其他错误。打开解析详情查看失败代码片段，修正后重新执行；不要只看任务整体“提交成功”。

空间质量检查

每轮任务完成后检查：

- 1. 文件解析成功率和部分成功数量。
- 2. 元数据覆盖率及未绑定对象。
- 3. 上下游链路是否出现断点、环线或孤立节点。
- 4. 解析版本、任务时间和来源文件是否可追溯。
- 5. 业务责任人抽查关键表、字段和报表链路。

覆盖率和图节点数量是质量线索，不是业务链路正确性的单一证明。

版本、编辑与回滚边界

自动解析结果应保留来源文件和任务记录。手工编辑只用于有明确证据的关系补充，不能用来掩盖采集或解析失败。变更前记录对象、方向、原因和责任人；发现错误时优先重新采集或重新解析，再撤销不正确的手工关系。

常见排查

现象	排查顺序
任务成功但没有图	检查资源是否入空间、命名空间是否绑定、解析报告是否有部分成功
表节点存在但字段链路为空	检查字段采集、脚本字段引用和强制绑定设置
同名表连错	检查命名空间、Schema 和 DBLink 映射
报表链路缺失	检查报表采集、数据连接和准备 SQL 解析引擎
增量任务重复解析	对比文件路径、名称、大小和内容，确认任务模式

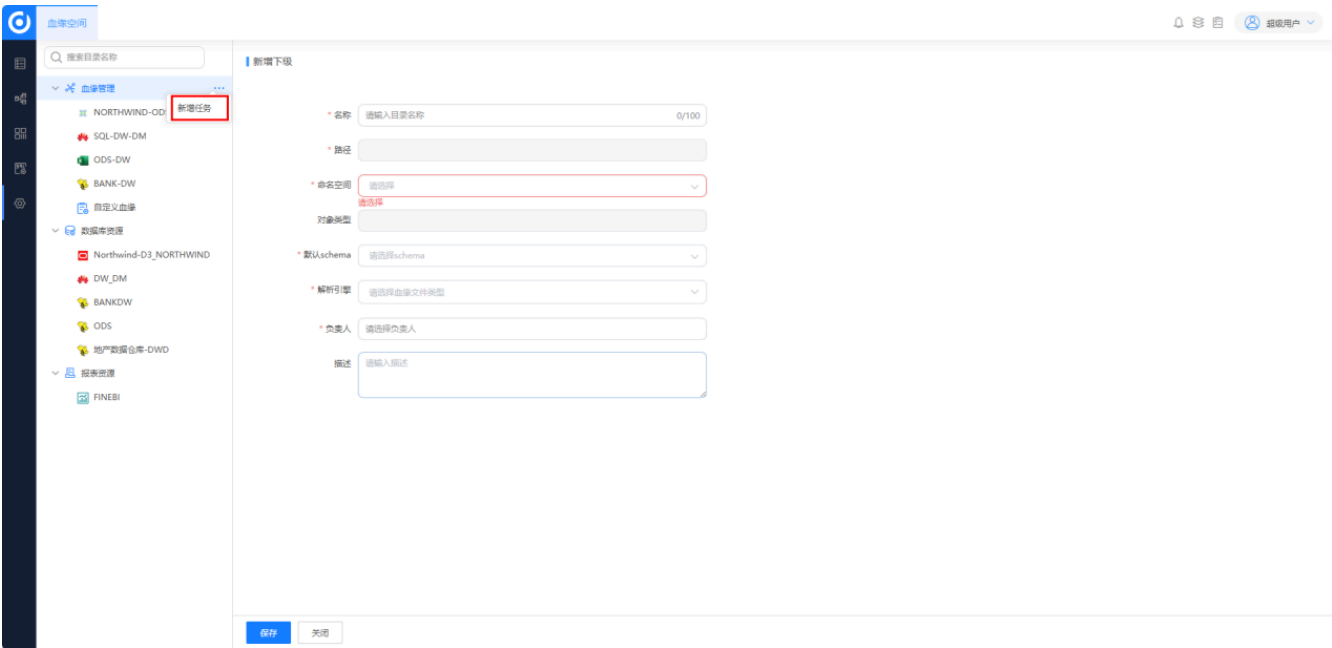
相关页面：[采集资源](#)、[血缘任务](#)、[查看和分析血缘](#)。

创建血缘任务并检查解析报告

血缘任务组织原文件、预处理、解析引擎及对象匹配。先准备元数据和命名空间，再用少量文件验证后扩大范围。

创建任务

- 1. 进入“系统管理 → 血缘管理 → 血缘空间 → 血缘管理”。
- 2. 在目标目录选择“新增任务”。
- 3. 按来源选择 Script 或 ETL。Excel 自定义血缘属于 ETL; Container of User-Defined-Object 是手工血缘编辑保存的不可解析类型。
- 4. 配置任务、引擎与命名空间，导入对应格式文件。
- 5. 检查预处理变量及数据连接，将每个连接映射到正确命名空间。



图示展示任务配置位置，任务类型和对象应使用实际业务值。

选择导入方式

方式	行为
全量导入	导入指定目录脚本，重复文件覆盖并重置为未解析；支持多层目录和不同层同名文件
增量导入	比较内容、大小、路径和文件名，相同跳过，不同更新

导入完成后检查文件数量、路径及状态；不要仅根据文件名判断是否重复。

设置元数据限制

默认不强制绑定元数据。开启强制绑定后，任意表或字段无法在命名空间找到对应元数据，都可能导致任务失败。因此先检查元数据采集、Schema、对象命名和连接映射。

存在任何未关联命名空间的数据连接会导致解析失败。跨系统 ETL 尤其应逐个检查源、目标连接。

执行与阅读报告

执行文件解析后，每次生成一个解析任务及起止时间、状态和统计。报告按成功、部分成功、语法错误、元数据缺失、其他错误分类。

1. 查看总量与各类结果，打开文件级错误明细。
2. 对部分成功文件查看高亮失败代码段。
3. 比较原文件和预处理后文件，判断变量及语法是否正确处理。
4. 修正引擎、连接、元数据或预处理后重新解析。
5. 在实际血缘图抽查关键节点、方向及字段关系。

成功率与元数据覆盖率分别反映解析处理和对象匹配，不应只追求成功文件数而忽略错误来源匹配。

相关文档：[命名空间与预处理](#)、[查看血缘](#)。

查看和分析血缘

本指南统一说明数据目录、血缘地图和应用场景中常用的血缘图操作，包括搜索、分析方向、层级、深度、筛选、导出、二级血缘、代码裁剪和差异比较。

阅读图形时先区分三个维度：**方向**是相对当前对象看上游、下游或全链路；**层级**是看系统/Schema、表等实体、字段等属性，或加工算子；**深度**是沿关系向外展开的步数。改变其中一项不会自动改变另外两项，屏幕上未显示的节点也不能直接判定为不存在。

血缘搜索

全局搜索

全局搜索是从血缘地图快速打开对象血缘的入口。可搜索 Schema、实体、属性和脚本文件：

- Schema、实体和属性支持多选，并可同时查看其全链路血缘。
- 脚本文件会在新的血缘视图中打开详细关系。
- 多选前确认对象所属系统和命名空间，避免把同名对象放入同一分析。

图内搜索

图内搜索只在当前已打开的血缘图中定位对象，可搜索 Schema、实体、属性或标签并高亮其位置。它不会扩大当前血缘深度，也不会加载尚未展示的对象。

上游 下游和全链路

当前实体是血缘图的分析基准。以当前实体为参照：

- **上游分析**展示产生或影响当前实体的对象。
- **下游分析**展示当前实体可能影响的对象。
- **全链路**展示与当前实体相连的全部上下游关系。

在图中切换分析方向后，应再次确认当前实体标识。多个对象同时作为当前实体时，导出和影响分析的范围也会随之变化。

血缘层级

层级控制血缘图的对象颗粒度。血缘地图默认可从平台级或 Schema 级开始，通过钻取进入实体级和属性级，也可以从属性级向上返回实体级或 Schema 级。

在本页中，Schema 是数据库组织表等对象的逻辑名称范围；“实体”主要指表、视图、报表或文件等可独立识别的对象；“属性”指对象中的字段等组成项；“算子”指加工过程中可解释的计算或转换步骤。这些名称描述图中的对象类型，不表示四种不同的授权产品。

层级	适合回答的问题
Schema 级	数据在系统或数据库区域之间如何流动
实体级	哪些表、视图、报表或文件存在依赖
属性级	某个字段或指标具体来自哪里、流向哪里
算子级	字段在 SQL 加工步骤中如何计算或传递

血缘深度

深度表示从当前实体向上游或下游展开多少组直接关联对象。深度越大，加载的对象和关系越多。建议先以较小深度定位关键路径，再逐步增加，避免大图中难以识别重点。

筛选过滤

筛选支持按实体类型或标签隐藏和显示对象，并支持单选、多选。被取消显示的项数会在筛选图标上提示。

例如链路为 $A \rightarrow B \rightarrow C$ ，隐藏中间实体 B 后，画面可显示为 $A \rightarrow C$ ；关闭筛选后恢复原始展示。筛选只改变当前视图，不会删除 B 或底层关系。

常见用途包括隐藏 ETL 临时表、仅查看核心表、按责任域或系统标签聚焦一段链路。

图形控制

控件	作用
SQL	在脚本文件血缘中显示或隐藏 SQL 窗口
放大和缩小	调整对象显示比例
全屏	扩大血缘图工作区域
显示	控制临时对象、中文名称、间接关系等内容
编辑	进入受控的血缘编辑状态

编辑状态还提供新建、撤销、重做、删除和保存。删除只允许作用于本次编辑新增的对象或连线，不能删除画布原有内容。

导出影响报告

影响报告默认以当前实体为基准，可以配置：

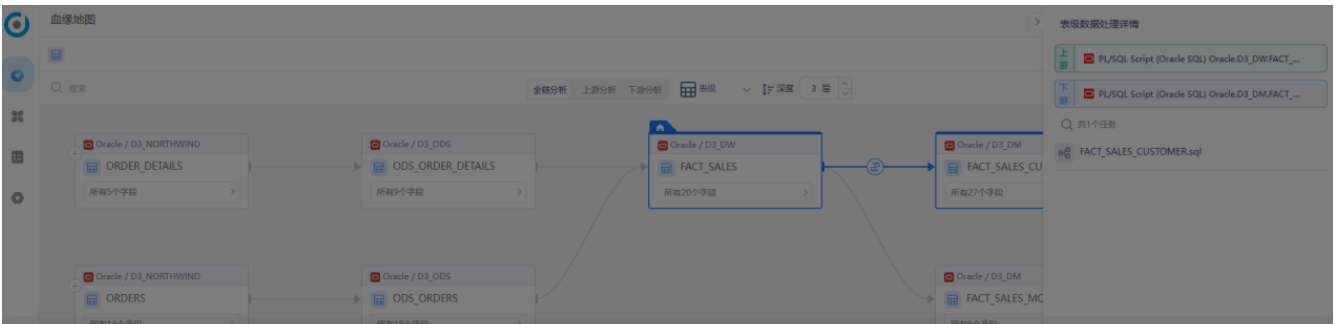
- 表级或字段级对象颗粒度。
- 上游、下游或全链路范围。
- 对象类型。
- 是否排除中间处理过程。
- 当前页面支持的导出格式。

报告可以包含因画面控制而未显示的对象，因此报告范围可能大于当前屏幕。也可以右键选择图中对象，通过“影响分析”生成该对象全部深度的影响报告。

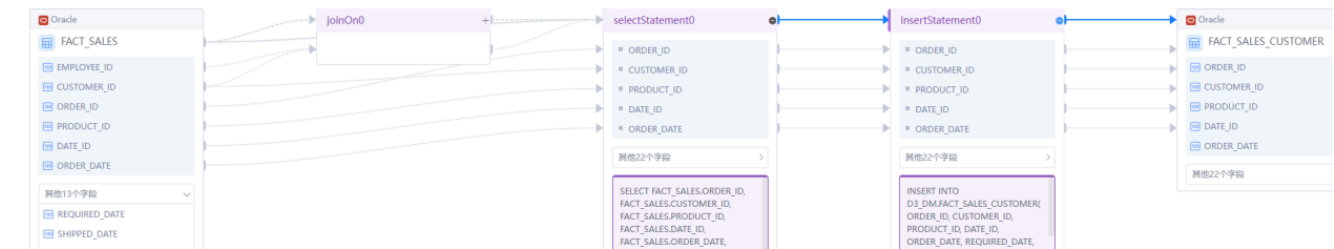
当图中存在多个当前实体时，系统会分别生成对应报告；没有明确当前实体时，范围可能按图中全部对象计算。导出前必须检查分析基准。

二级血缘和算子血缘

“二级血缘”是在当前对象关系图之下继续展开的处理细节视图，不是第二套独立血缘数据。表、字段和算子支持进入二级血缘；其中“算子血缘”把 SQL 或脚本中的计算、转换步骤与输入、输出对象关联，用于解释字段如何产生或传递。字段旁的 **fx** 标识表示存在计算表达式，可以高亮某一算子的上下游，并进入算子级关系。



加工过程图示



二级血缘用于解释处理过程，不应脱离完整 SQL、变量和调度上下文单独作为业务结论。

代码裁剪

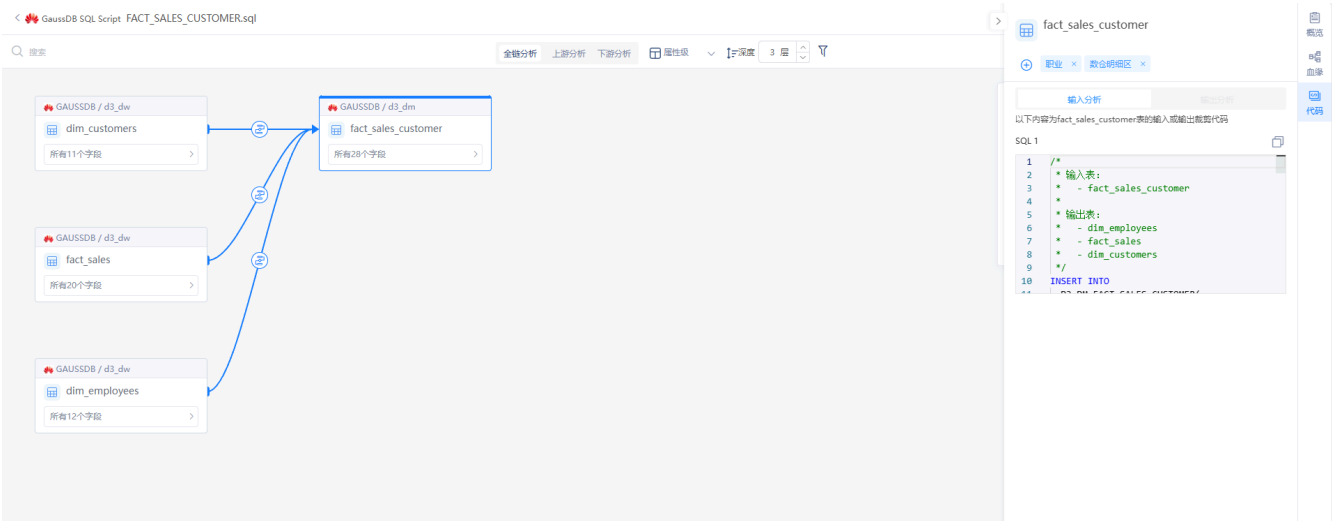
代码裁剪从完整脚本中提取与选中实体或属性相关的处理片段，便于定位该对象的输入或后续使用。这里的“裁剪”只改变分析时展示的代码范围，不表示修改或删减源脚本。

查看实体上游处理逻辑

- 1. 在血缘图中选择目标实体。
- 2. 打开对象的右键菜单并选择“代码裁剪”。
- 3. 选择输入分析。
- 4. 查看向该实体写入数据的相关代码片段。

查看字段下游使用逻辑

- 1. 展开实体并选择目标字段或指标。
- 2. 打开“代码裁剪”。
- 3. 选择输出分析。
- 4. 查看后续视图、计算或连接中使用该字段的代码片段。



代码裁剪适合口径穿透和跨步骤逻辑整合。结果是聚焦片段，不等同于完整可执行脚本。

比较文件和血缘差异

代码差异

在血缘空间打开血缘文件的比较功能，可以逐行查看两个文件修订之间的变化。页面使用语法着色和差异高亮帮助识别字段、表达式和结构变化。

版本比对

V2

V1

```
1 CREATE GLOBAL TEMPORARY TABLE tmp_salary (employee_id, total_salary)
2 DELETE
3 ROWS;
4 INSERT INTO
5 tmp_salary (employee_id, total_salary)
6 SELECT
7 s.employee_id,
8 SUM(s.salary) AS total_salary
9 FROM
10 salary s
11 GROUP BY
12 s.employee_id;
13 CREATE
14 OR REPLACE VIEW employee_income AS
15 SELECT
16 e.employee_id,
17 e.lastname,
18 e.firstname,
19 e.gender,
20 e.date_of_birth,
21 e.date_of_hire,
22 d.department_name,
23 ts.total_salary,
24 b.bonus
25 FROM
```

```
1 CREATE GLOBAL TEMPORARY TABLE tmp_salary (employee_id, total_salary)
2 DELETE
3 ROWS;
4 INSERT INTO
5 tmp_salary (employee_id, total_salary)
6 SELECT
7 s.employee_id,
8 SUM(s.salary) AS total_salary
9 FROM
10 salary s
11 GROUP BY
12 s.employee_id;
13 CREATE
14 OR REPLACE VIEW employee_income AS
15 SELECT
16 e.employee_id,
17+ e.name,
18 e.gender,
19 e.date_of_birth,
20 e.date_of_hire,
21 d.department_name,
22 ts.total_salary,
23 b.bonus
24 FROM
```

血缘差异

血缘差异根据两个解析结果比较节点和关系的变化。页面可用颜色区分新增、修改和删除对象，并通过切换控件查看两侧状态。

比较前确认两次解析使用相同的采集范围、命名空间和解析规则，避免把配置变化误判为业务变化。

结果验证

- 当前实体、所属系统和命名空间正确。
- 上下游方向与实际数据处理方向一致。
- 层级和深度足以回答问题，没有因筛选遗漏关键对象。
- 二级血缘和代码片段与完整脚本可以相互印证。
- 影响报告的基准、方向、颗粒度和范围正确。
- 差异比较两侧的采集与解析条件具有可比性。

相关文档

- [数据血缘探索](#)
- [应用场景](#)
- [在线 SQL 血缘](#)
- [血缘系统管理](#)

补充并管理手工血缘

手工血缘用于有明确事实依据的小范围断链补充。对持续出现的同类缺失，优先修复采集和解析，不长期依靠人工猜测关系。

编辑操作

1. 在“血缘地图”进入属性级血缘，从工具栏选择“编辑”。
2. 根据需要新建实体，核对对象身份。
3. 钉住起始和目标字段，从起始字段的加号拖至目标字段建立连接。
4. 需要取消连线时单击空白处；用撤销、重做调整本次操作。
5. 删除只作用于本次新增实体或绿色连线，不能删除原画布内容。
6. 检查方向、对象和依据后保存。

保存会生成带时间戳和 ULE 扩展名的自定义血缘文件，并退出编辑。退出、切换或刷新页面会放弃未保存编辑。

保存后的限制

已保存手工链路可以查看或删除，不能再次编辑修改。更正前先记录原关系和理由，再按允许的管理方式处理，不能假定存在对已保存文件的撤销功能。

后台管理

在“系统管理 → 血缘空间 → 血缘管理 → 自定义血缘”查看生成文件，也可从图中实体间的任务列表查看。删除文件会影响其关系，应先确认目标文件及业务影响。

最终检查源目标方向、字段粒度及关联元数据，并在治理记录中保留来源、维护人和原因。

配置命名空间、DBLink 与预处理

命名空间区分不同系统或环境中的同名对象；预处理把原始代码中的变量、连接和不规范内容转换为可解析输入。两者共同影响血缘是否关联到正确对象。

规划命名空间

进入“系统管理 → 血缘管理 → 命名空间”，创建或编辑名称及对应属性。名称唯一，设计时应明确技术来源、环境和生命周期。

同一数据库中执行的 SQL 通常与其元数据使用相同命名空间。跨系统 ETL 需要将多个数据连接分别映射到源、目标命名空间。

默认 Schema 用于表名前缺少 Schema 的情况。填写错误会使解析节点落到错误路径，不能以补填任意 Schema 消除缺失提示。

配置 DBLink

在 DBLink 映射管理中查看采集到的 DBLink，必要时手工创建，核对以下内容：

项目	含义
DBLink 名称	代码引用的连接关键字，例如示例 customer@remote 中的 remote
默认 Schema	经该连接访问表但未指定 Schema 时的补充值
创建空间	DBLink 所在数据库对应的命名空间
指向空间	DBLink 实际连接的数据库对应命名空间

解析前完成映射，再用跨库样本检查是否连接到正确来源。

创建并应用预处理

1. 进入“预处理管理”，选择 JavaScript、正则表达式或大模型方式。
2. 为特定数据库和代码场景定义处理内容；使用大模型前先配置相应接口。
3. 在血缘空间任务的“预处理”中选择适用规则。
4. 检查自动提取的变量与数据连接，必要时批量导出、维护、导入。
5. 对少量文件比较处理前后内容，确认仅解决解析问题而未改变业务含义。
6. 解析后检查文件报告和链路。

大模型可辅助从 Java、Python 中提取 SQL 或修正语法，但应复核输出。保留原文件，不把经过推断改写的语句当作业务系统原始执行事实。

排查断链

优先检查元数据是否采集、连接是否全部映射、默认 Schema 是否正确、DBLink 源目标是否颠倒、变量是否被替换，以及预处理是否遗漏关键语句。修复后重新解析并核对结果。

查询血缘图谱、加工逻辑与影响范围

本页整理原接口手册中有参数表和响应示例的查询接口，供第三方集成时核对。认证和服务入口遵循[集成约定](#)。接口方法均为 POST，下列路径追加在服务的 /sqlink 基础路径后。

文中对象 ID、标签和示例值需要替换为目标环境的实际标识。不要把显示名称、物理表名或另一个环境的 ID 直接当作同一对象。

“血缘子图”是以指定对象为起点，按方向、深度和过滤条件截取的一部分血缘节点与关系，不是整个 SQLink 血缘库；“加工逻辑”是连接输入与输出的转换语句或处理描述；“影响范围”是依据当前已采集、已解析关系得到的潜在关联对象，不等于已确认的实际业务损失。使用接口结果前应核对这三个口径。

查询血缘子图

需要跨模块组合元数据、资产和血缘时，可先查看 [DAM REST API 场景指南](#)。该指南按服务职责筛选接口，并说明与旧接口手册的枚举差异；实际调用采用目标部署的契约。

路径：/lineageGraph/query。参数放在 JSON Body 中。

参数	类型	必填	含义及原接口说明的默认值
id	String	是	目标节点 ID
queryType	String	否	SCHEMA、TABLE、COLUMN；默认 TABLE
depth	Integer	否	查询深度；默认 3
direction	String	否	FORWARD 下游、BACKWARD 上游、ALL 全链路；默认 ALL
hasTemp	Boolean	否	是否显示临时表；默认 false
tags	Array	否	标签过滤；原示例使用标签名称列表，与添加标签接口的 tagIds 不同

调用时建议显式填写影响范围的参数，避免默认值与使用目标不一致。以下为请求结构示例，example-node-id 不是可直接调用的真实标识：

```
{
  "id": "example-node-id",
  "queryType": "TABLE",
  "depth": 3,
  "direction": "ALL",
  "hasTemp": false,
  "tags": []
}
```

```
{
  "id": "example-node-id",
  "queryType": "TABLE",
  "depth": 2,
  "direction": "FORWARD",
  "hasTemp": false
}
```

响应示例包含 nodes 和 relationships。节点记录中有 id、name、parentId、type 和 properties；关系通过 srcId、dstId 关联节点，并提供 lineageId 及关系属性。

检查返回结果时，先核对起始对象、方向、深度和过滤条件，再检查节点与关系能否对应。某对象未出现，可能受本次深度或过滤影响，不能直接判定全平台没有此对象的血缘。

获取加工逻辑

路径：/lineageGraph/getTransform。

- Query 中的 isGraph 为可选 Boolean，原说明默认 true，用于图内模式。
- Body 使用图查询条件，并指定目标 id。原说明也提到 queryIds，但未完整给出它的参数结构；批量调用前应以目标接口定义为准。
- 返回加工逻辑列表，示例字段包括 startIds、endId、transforms、lineageName、lineageId 和 lineageTransformationType。

transforms 中的转换语句用于解释加工关系。将其与源脚本、输入对象和输出对象相互核对，不把聚焦片段直接当作可独立运行的完整作业。界面操作见[代码裁剪](#)。

查询影响范围

路径：/lineageGraph/getImpact。参数放在 JSON Body 中。

参数	类型	必填	含义及原接口说明的默认值
id	String	是	分析基准节点 ID
direction	String	否	FORWARD 或 BACKWARD，默认 FORWARD；不要直接套用图查询的所有方向枚举
depth	Integer	否	影响分析深度，默认 3
resourceTypes	Array	否	可过滤 TABLE、REPORT、COLUMN；需要限制时显式传入数组

响应为影响对象列表，示例字段有 id、name、alias、type、dataArea 和 direction。原参数表对不限类型的描述与数组类型未完全对应，因此不要自行把字符串 ALL 填入数组参数；按目标接口定义处理不限范围的调用。

影响列表提供变更影响线索，不自动通知责任人或修改下游资产。结合[应用场景](#)检查业务影响。

查询命名空间统计

路径：/dataArea/getMetadataCoverage。Query 参数 namespaceId 必填，类型为 Long。

响应中的统计字段包括：

粒度	血缘侧计数	元数据侧计数
Schema	schemaInLineage	schemaInMetadata
表	tableInLineage	tableInMetadata
字段	columnInLineage	columnInMetadata

先确认命名空间和统计粒度，再与[资源同步和采集范围](#)核对。上述计数不能单独证明每个对象已——正确匹配；覆盖率使用页面或目标接口明确提供的口径，不根据字段名称自行拼出百分比。

根据标签查关联对象

路径：/tag/getByTag。Query 参数 tagId 必填，类型为 Long。使用实际标签 ID，不以标签名称代替。结果用于查找已关联对象；查询本身不会为对象添加标签，也不会发起标签扩散。

响应与异常检查

HTTP 状态	原接口说明	调用方的检查
200	请求成功	继续核对业务内容、范围及结果字段
400	参数错误	参数位置、必填项、类型、枚举及对象标识
401	认证失败	当前认证信息与访问权限，不重复提交无效凭据
404	路径或资源不存在	基础路径、对象是否存在及访问范围
500	服务异常	保存时间、接口及必要的错误摘要，由服务管理员检查

接口联调记录应包含请求范围、对象标识、响应状态和业务结果，不记录真实 Token 或密钥。页面能打开、HTTP 返回成功和业务关系正确需要分别核对。

在线 SQL 血缘

在线 SQL 血缘面向数据开发和分析人员，用于在不创建正式解析任务的情况下快速检查一段 SQL 的输入、输出、中间过程和字段关系。在线结果适合开发核对和问题分析；需要持续管理的代码应进入正式采集与解析流程。

使用前提

- 已取得待分析 SQL 的合法使用权限，并已移除密码、令牌等敏感信息。
- 已知 SQL 所属数据库类型或适用的解析引擎。
- 如需关联 DAM 元数据，已确认相应命名空间和默认 Schema。

首次操作可直接使用[快速开始的教学样例](#)，先识别输入、输出和字段，再分析真实业务代码。

解析 SQL

1. 进入在线 SQL 血缘页面。
2. 选择 SQL 类型或与语法匹配的解析引擎。
3. 将 SQL 输入编辑区。
4. 如需整理缩进和换行，执行美化 SQL。
5. 打开解析设置，按实际需要配置间接关系和元数据关联选项。
6. 执行解析。
7. 在结果区切换表级或字段级视图并核对关系。

The screenshot displays the Online SQL Lineage tool interface. On the left, the SQL editor shows two statements. Statement 1 is an INSERT into dwd.order_summary, and Statement 2 is a UNION query inserting into dws.final_report. The right side shows a dependency graph with nodes for ods.orders, ods.customer, dwd.order_summary, and dws.final_report, connected by arrows indicating data flow. The interface includes a top navigation bar with 'SQL血缘' and '解析设置' tabs, and a bottom status bar with a '登录用户' button.

解析设置

设置	作用	使用建议
显示间接关系	展示通过中间表达式或过程推导的关系	排查复杂转换时启用；阅读主链路时可关闭
强制关联元数据	将解析节点绑定到已采集的元数据	仅在命名空间和 Schema 明确时使用
命名空间	限定对象所属的逻辑环境	必须与目标元数据的规划一致
默认 Schema	为 SQL 中未写 Schema 的对象补充上下文	按 SQL 实际执行环境填写，不作猜测

强制关联元数据时，如果命名空间或默认 Schema 配置错误，可能关联到错误对象或产生无法匹配的节点。首次使用应选择一段已知结果的 SQL 做小范围验证。

阅读解析结果

- 蓝色节点表示输入对象。
- 绿色节点表示输出对象。
- 黄色节点表示解析过程中形成的中间对象或过程。
- 实线表示直接关系。
- 虚线表示间接关系。

可在表级和字段级之间切换。表级用于快速理解整体输入输出，字段级用于核对列的来源和转换。完整链路开关用于控制是否展示通过中间节点展开的关系；复杂 SQL 下应结合筛选和缩放逐段检查。

运行历史

运行历史用于回看已执行的在线解析。打开历史记录后，核对 SQL、解析类型、执行时间和结果是否属于当前任务。历史记录可能包含业务代码，查看和共享仍受数据权限与保密要求约束。

结果验证

- 输入和输出对象与 SQL 的 FROM、JOIN、INSERT 或创建语句一致。
- 字段映射能解释别名、表达式和聚合结果。
- 直接关系与间接关系的显示符合当前分析目的。
- 强制关联后的对象属于正确命名空间和 Schema。
- 解析警告、未识别语句或缺失对象已经记录。

常见问题

解析失败

检查 SQL 类型、语法完整性、数据库特有语法和变量。若代码依赖运行时替换变量，应先接受控样本完成预处理，再解析可执行形态的 SQL。

节点未关联到元数据

检查元数据是否已采集，以及命名空间、默认 Schema 和对象大小写是否一致。不要为了得到连线而关联到名称相同但来源不同的对象。

解析结果过于复杂

先切换到表级视图，再关闭完整链路或间接关系，定位主输入和主输出后逐步展开。

相关文档

- [查看和分析血缘](#)
- [SQLink 快速开始](#)
- [血缘系统管理](#)