

2021



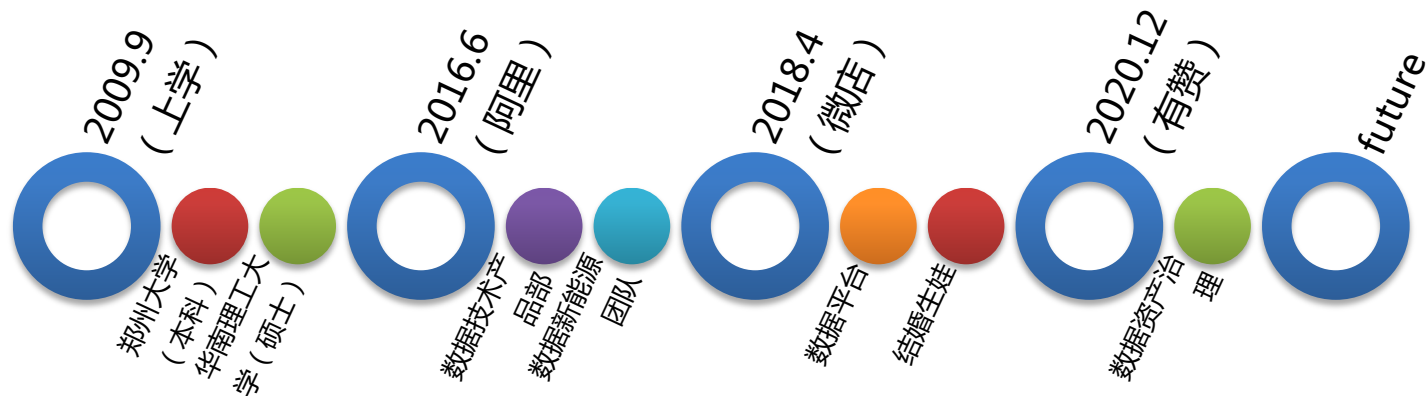
| DataFunSummit

有赞数据地图 实践

有赞 数据地图负责人：何会会



自我介绍



- ✓ 勇于试错、敢想敢做
- ✓ 总结过往、做好当下、不畏将来



目录

CONTENTS

- 数据地图背景
- 数据地图概述
- 数据地图实践
- 总结展望



01

数据地图背景



数据地图产生的背景

数据地图可以解决那些问题



数据地图背景

- ◆ 数据的流转链路不清晰
- ◆ 找不到想要的数据
- ◆ 无法高效的管理数据
- ◆ 故障排查效率低
- ◆ 链路优化困难



数据相关的工作过程

数据采集

数据开发

数据管理

数据搜索

数据分析

数据挖掘

故障排查

链路优化



■ 数据地图的目标

数据地图力图解决以下几个问题：

- ✓ 高效的找到自己想要的数据库
- ✓ 方便的查看多种多样的数据血缘信息
- ✓ 高效的对数据库进行管理
- ✓ 高效的应对故障（排查、影响面&恢复时间预估）
- ✓ 能够根据不同需求场景看到不同的链路视角



02

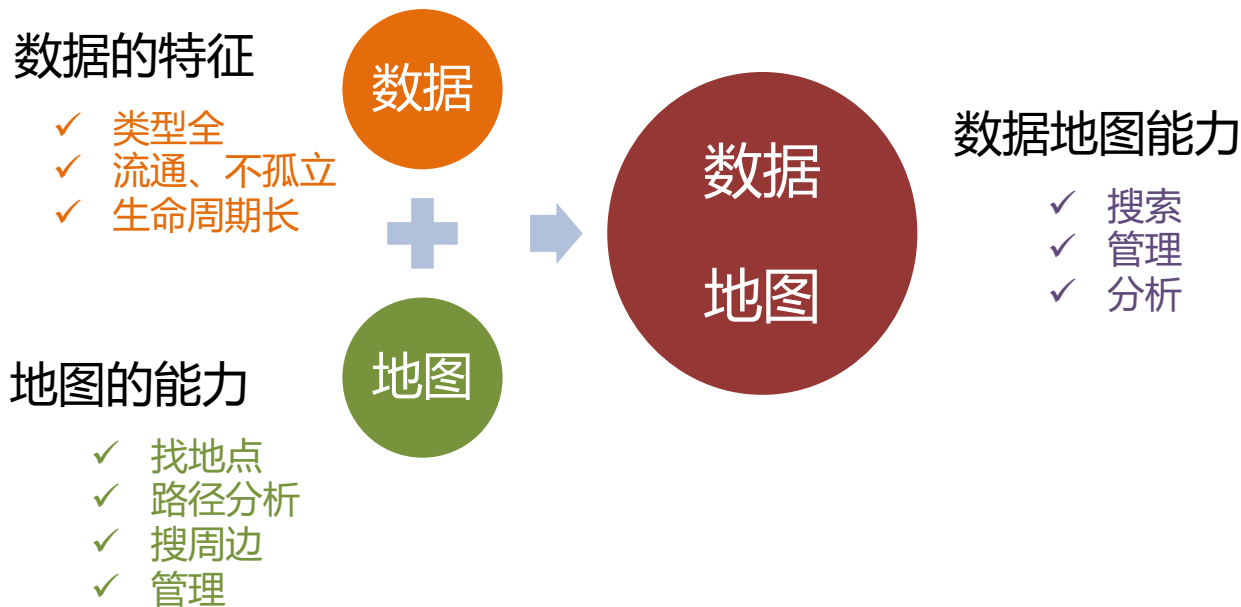
数据地图概述



数据地图概述



数据地图概述



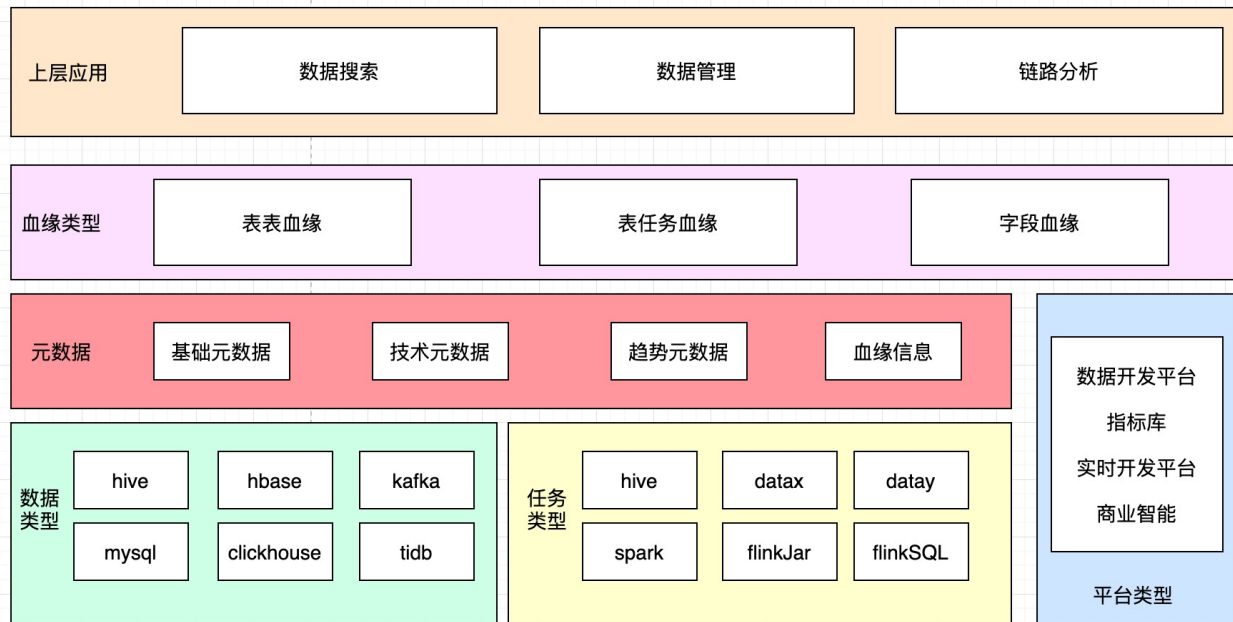
03

数据地图实践

- ✓ 数据全链路
- ✓ 数据搜索
- ✓ 数据管理
- ✓ 数据链路分析



实践-数据全链路



数据全链路：

- ✓ 数据类型全
- ✓ 任务类型全
- ✓ 平台类型全
- ✓ 元数据类型全
- ✓ 血缘类型全

对数据进行抽象成表和任务进行统一管理，完成了从业务到业务的闭环



实践-数据搜索

搜索的目标：

- ✓ 找数据更精准
 - ✓ 搜索结果更匹配
 - ✓ 搜索结果进行打分排序
- ✓ 找数据更容易
 - ✓ 从业务的角度搜数据



匹配内容：

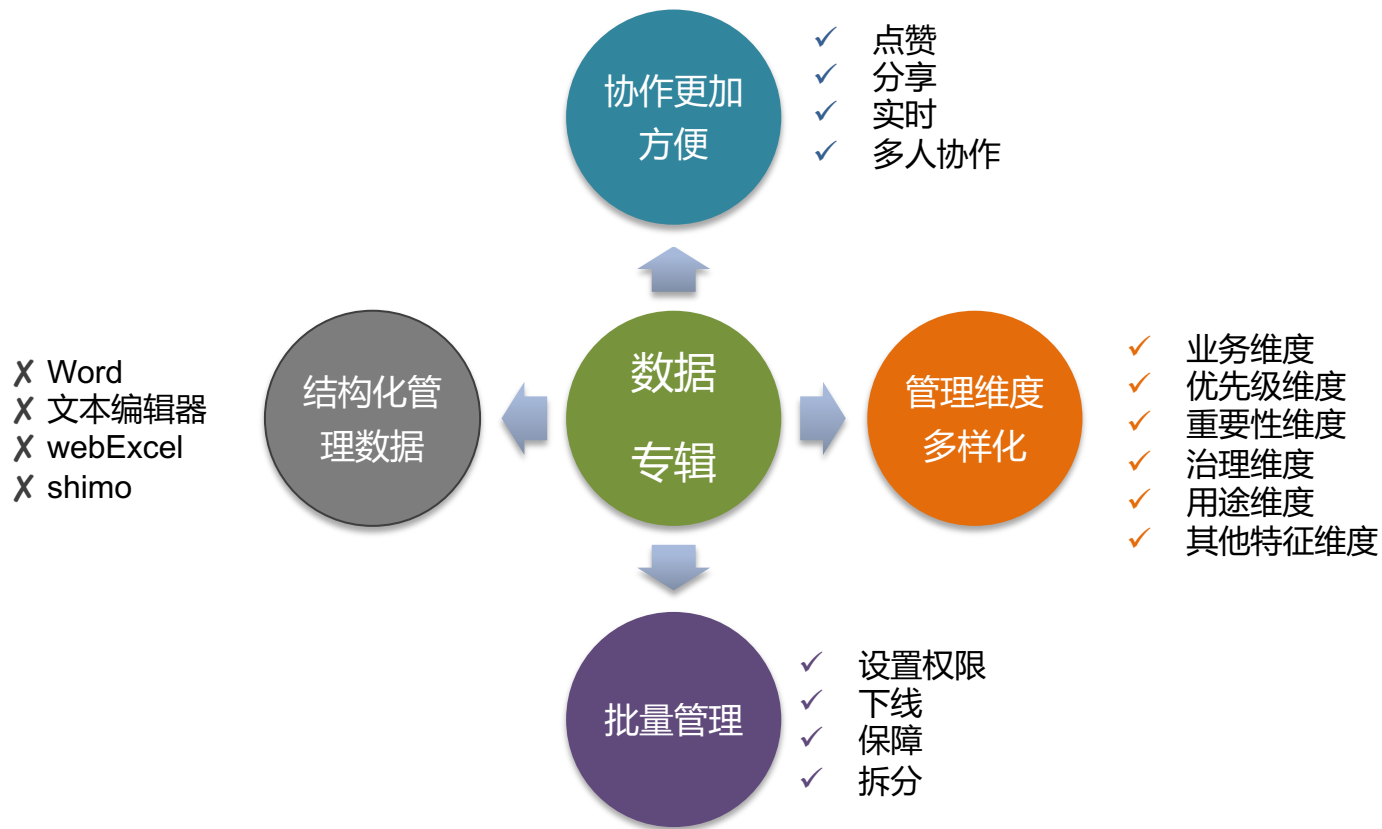
- ✓ 文本匹配
- ✓ 标签匹配
- ✓ 业务指标关联匹配
- ✓ 文档匹配
- ✓ 报表匹配

结果打分影响因素：

- ✓ 加分项
 - ✓ 当前owner
 - ✓ 下游数
 - ✓ 质量分
 - ✓ 访问次数
 - ✓ 公共层的表
- ✓ 减分项
 - ✓ 设置了替换表
 - ✓ 临时表



实践-数据管理



实践-血缘查看



- ✓ 表表血缘
- ✓ 表任务血缘
- ✓ 字段血缘

优化体验

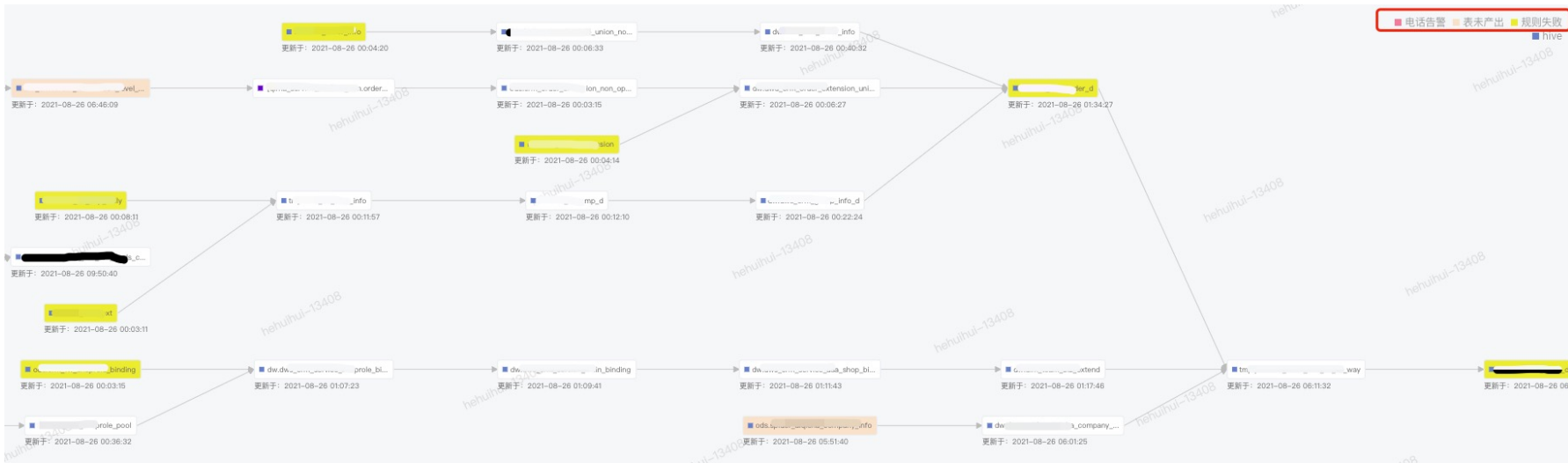
- 最上游&最下游
- 聚合查看
- 节点搜索
- 节点排序



DataFunSummit

实践-异常分析

当目标表出现异常时，需要排查原因时，就可以用到这个分析场景。



■ 血缘汇总信息

直接上游: 14

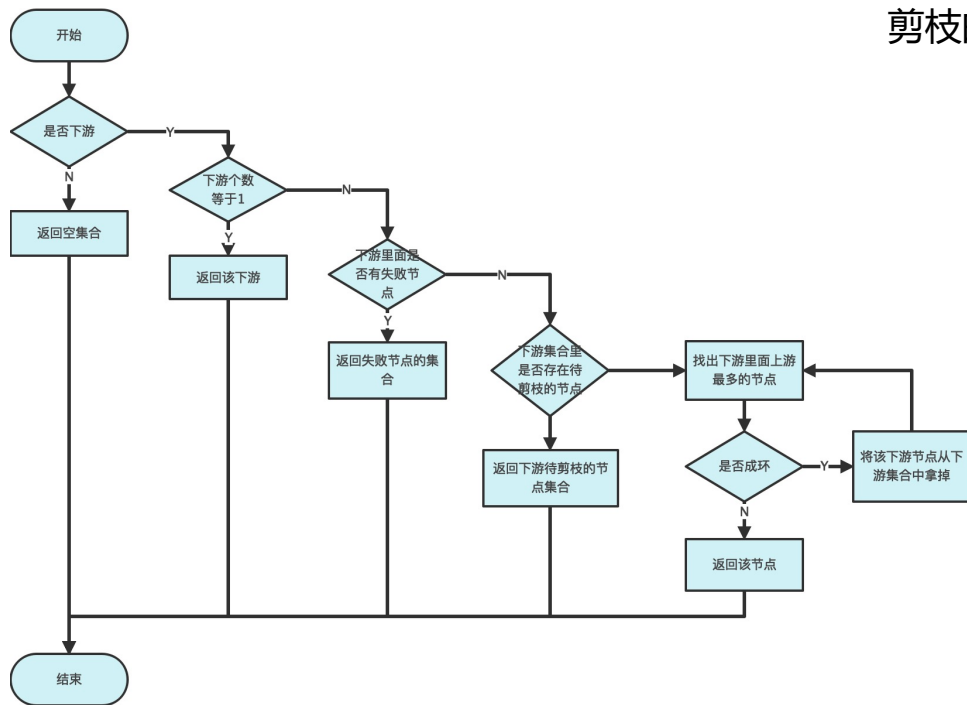
直接下游: 1

总上游: 1020

总下游: 22387

- ① 向上溯源，找到所有异常的表
- ② 以有异常的表为源头，经过**剪枝优化**，将相对简单的路径展示出来

异常分析剪枝算法



剪枝的目的：减少图中的非必要节点和边的数量

优化剪枝的关键步骤：

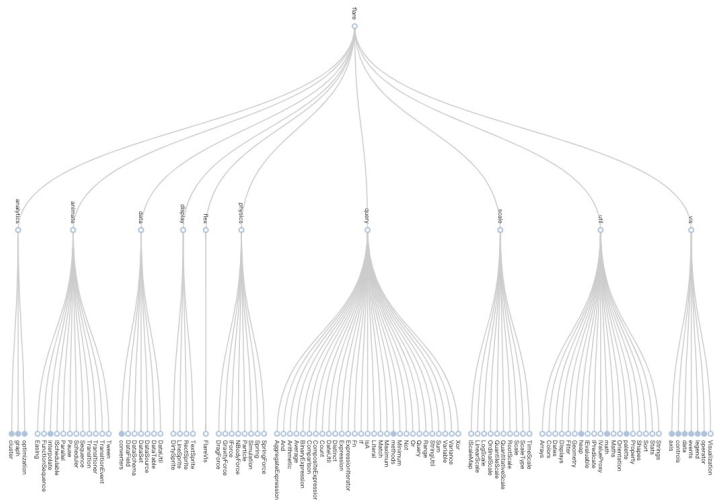
- ✓ 剪枝的起点
- ✓ 下游选择策略
 - ✓ 上游个数最多的节点（当前策略）
 - ✓ 最靠近目标表的（可尝试策略）



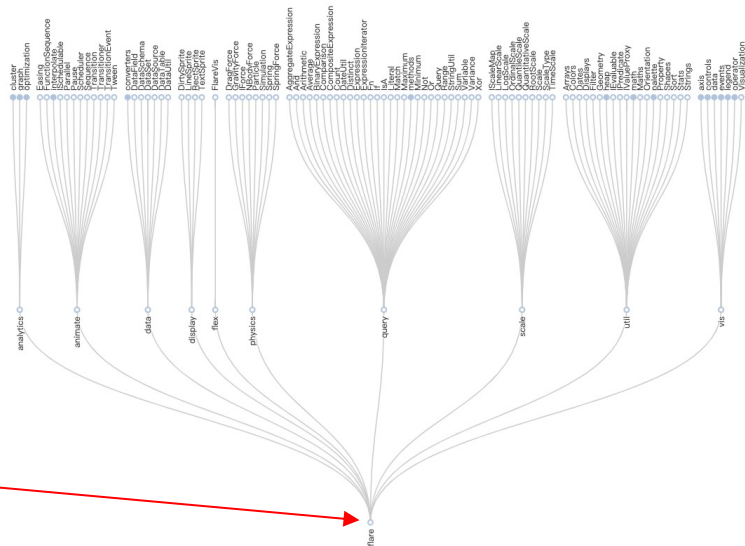
实践-影响分析&产出时间预估

当核心表出现故障的时候，需要使用这个分析场景，核心流程是：

- ✓ 向下评估影响面
- ✓ 向上预估产出时间



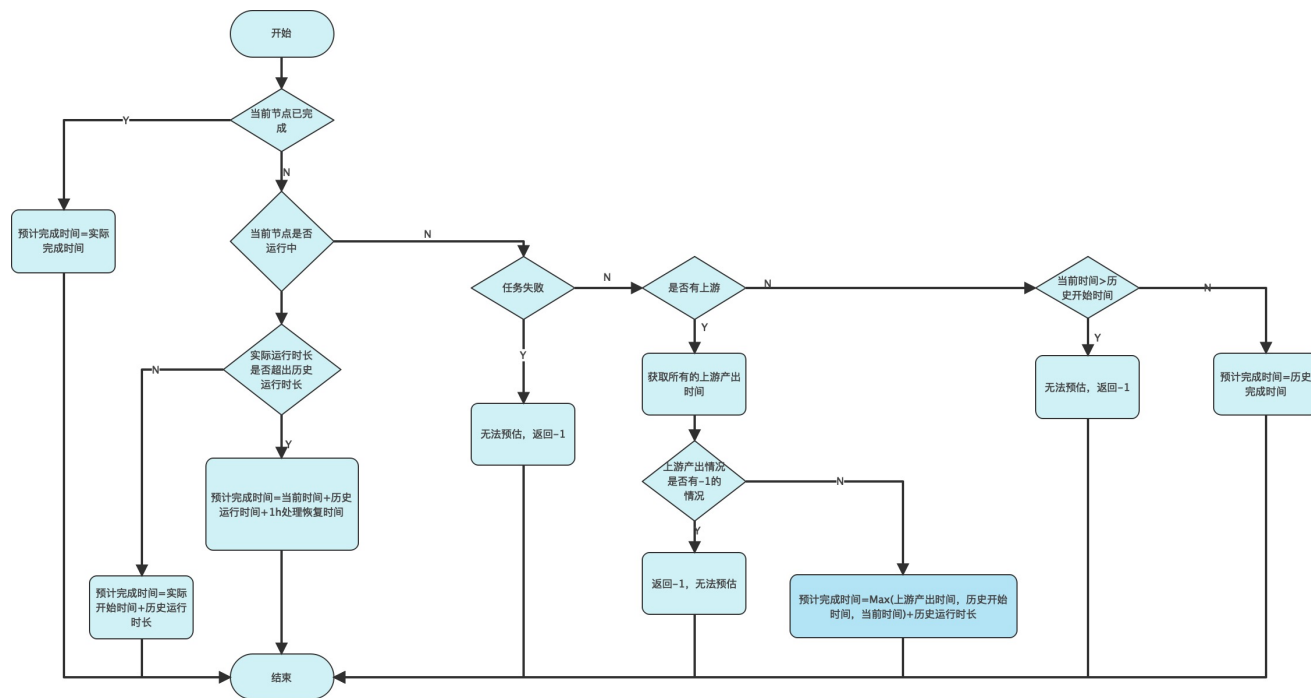
向下评估影响面



向上预估产出时间



预估产出时间算法



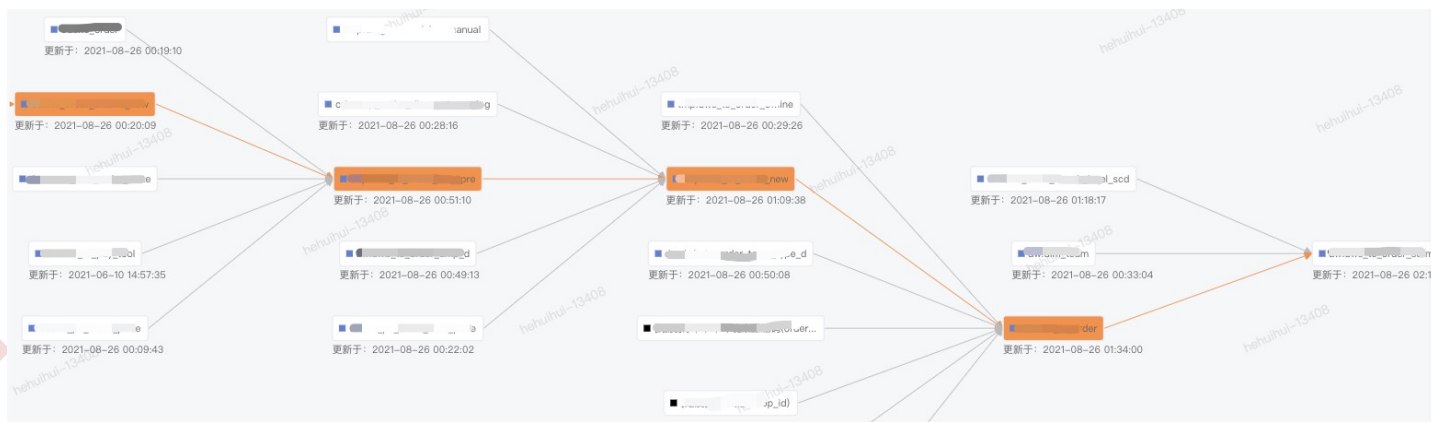
历史运行时长取的是最近7天的中位数



实践-链路优化

什么场景需要链路优化？

- ✓ 成本太大
- ✓ 链路太长
- ✓ 产出时间太晚



优化表产出时间场景：

- ✓ 关键路径
- ✓ 表任务血缘看任务启动时间是否合理
- ✓ 表是否可替换（根据字段血缘）



■ 实践-数据监控保障



04

总结与展望



- ✓ 总结
- ✓ 底层的存储方式重构
- ✓ 更多的场景支持
- ✓ 模型可视化



总结

- ✓ 更加流畅
- ✓ 接口RT<1S

- ✓ 提效1~3H



- ✓ UV:90->130
- ✓ PV:2K->3.5K

- ✓ 数据类型：29
- ✓ 任务类型：16
- ✓ 平台数：4+



■ 底层存储方式重构



血缘关系的底层存储从关系型数据库使用图数据库来重构

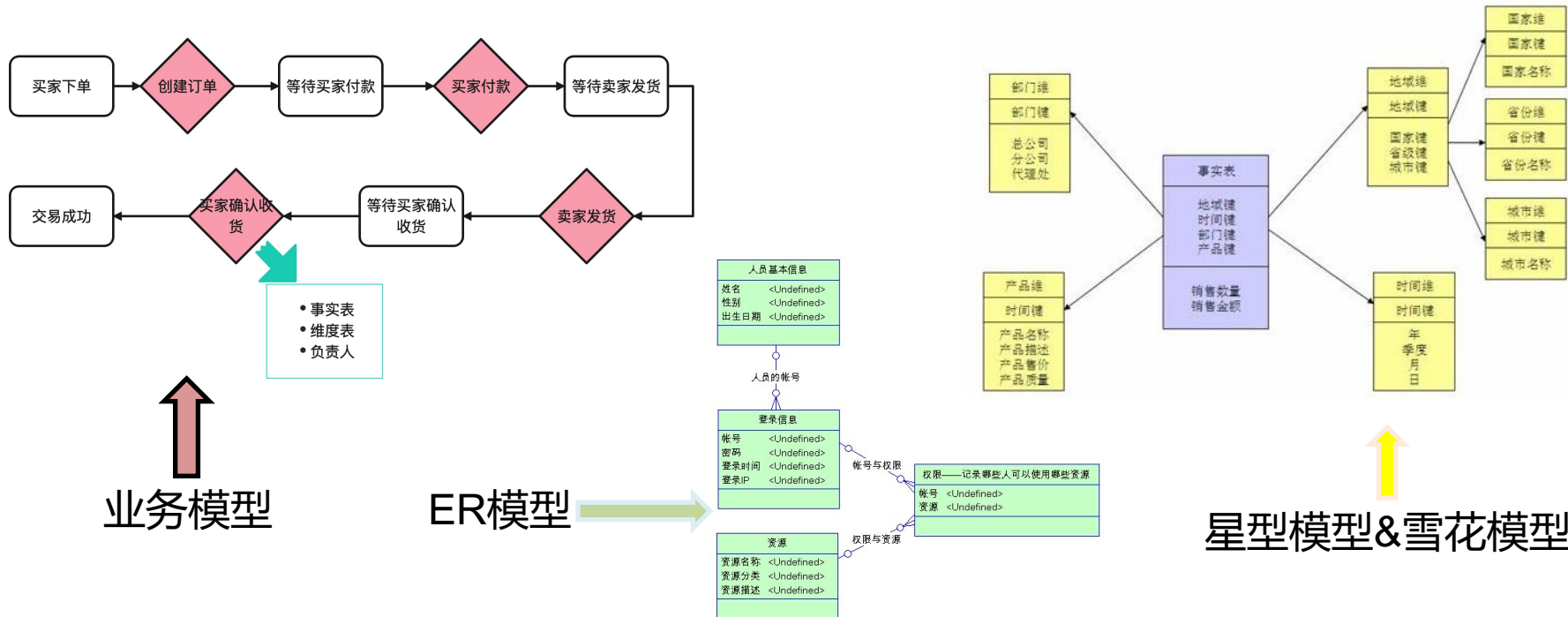


更多场景支持



模型可视化

支持更加丰富的模型可视化组件，将数据模型和业务模型直观呈现出来，更加容易和方便理解数据和业务。



2021



关注 “有赞coder” 公众号
每周都有干货



关注我，一起学习



Ending