

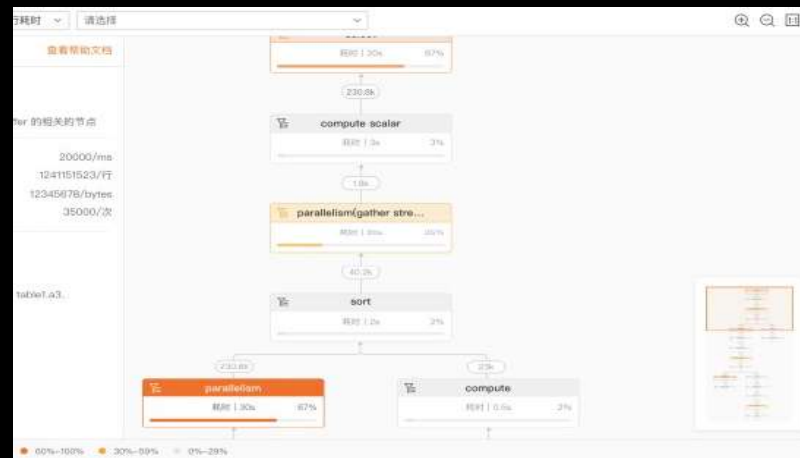
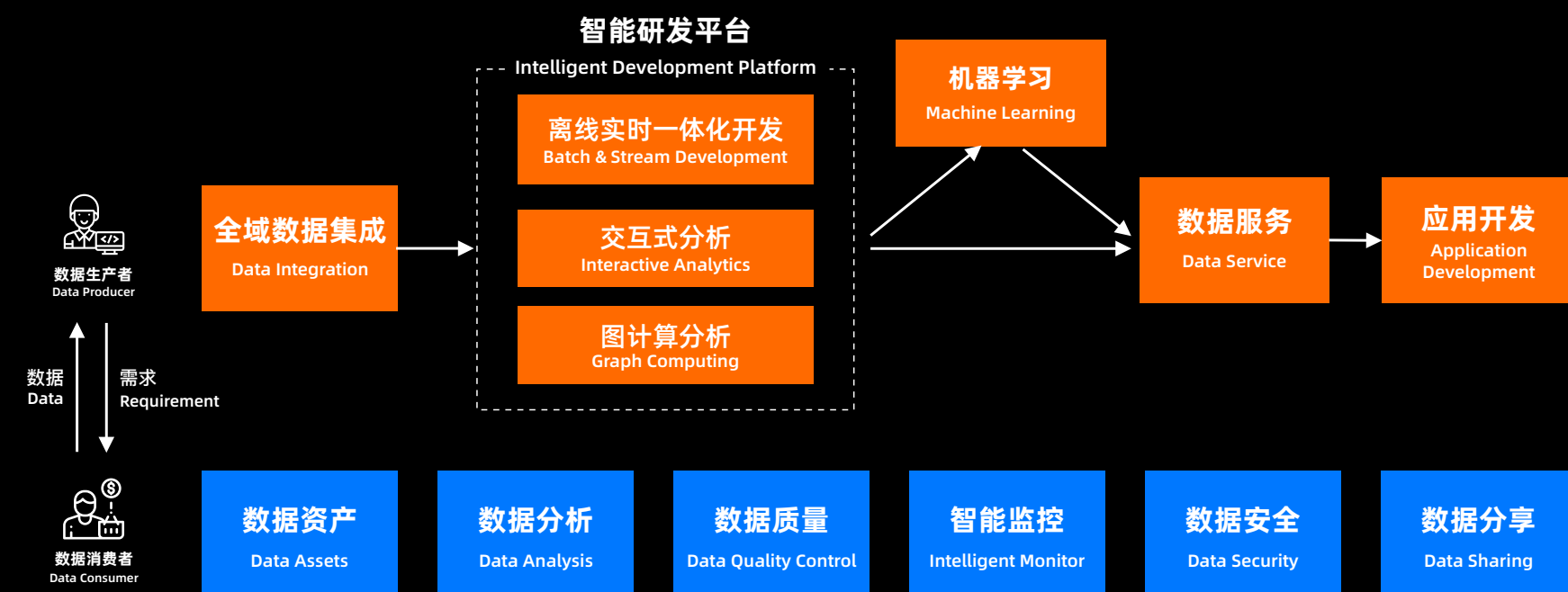
第十六届 D2 前端技术论坛

千万级节点网络下的可视分析技术

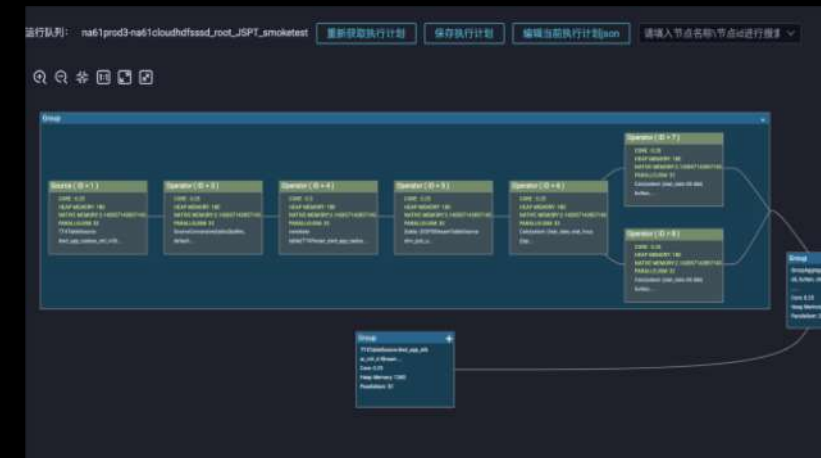
阿里云智能-计算平台事业部-DataWorks-刘东奇

DataWorks: 全域智能大数据平台

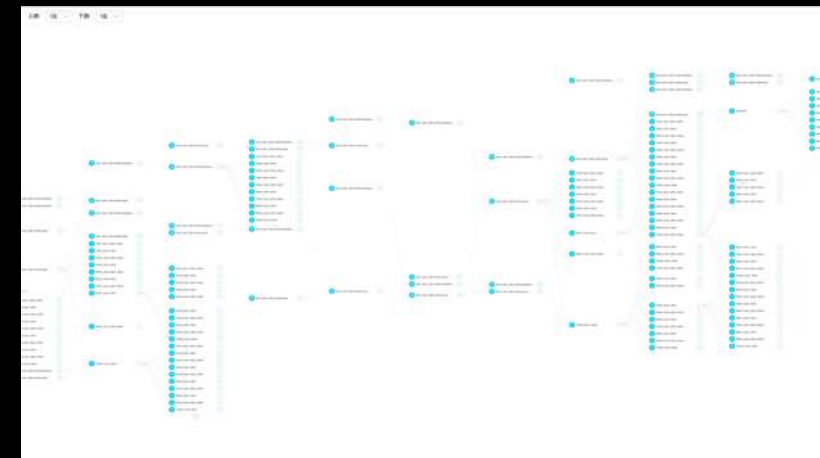
DataWorks: All-round Intelligent BigData Platform



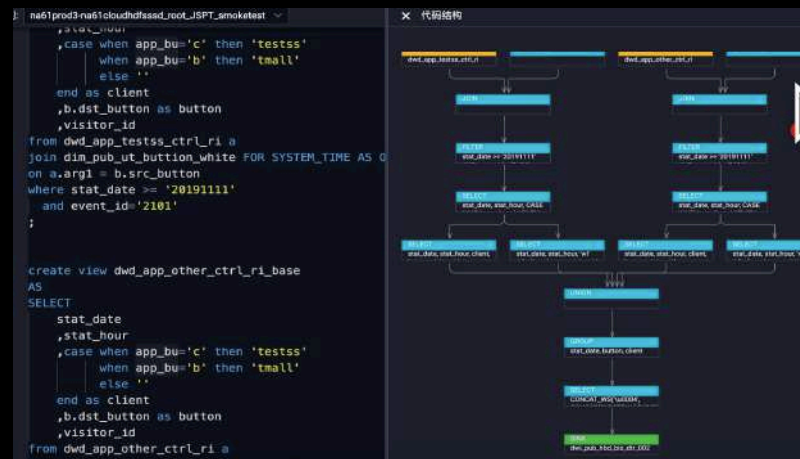
执行计划



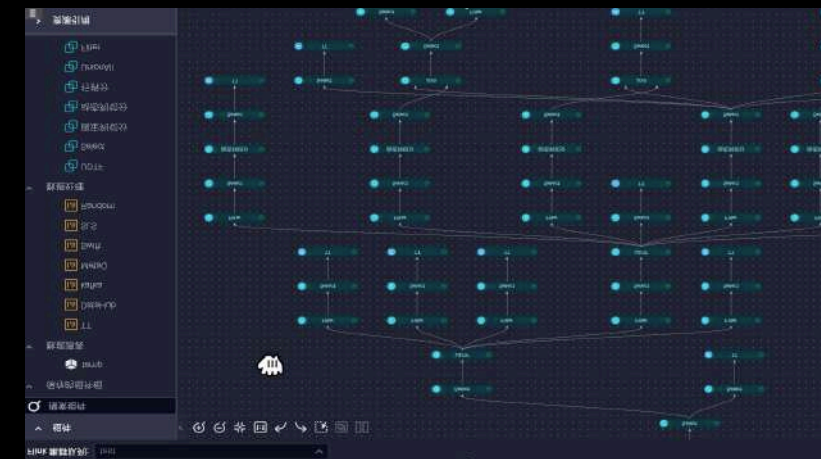
算子血缘



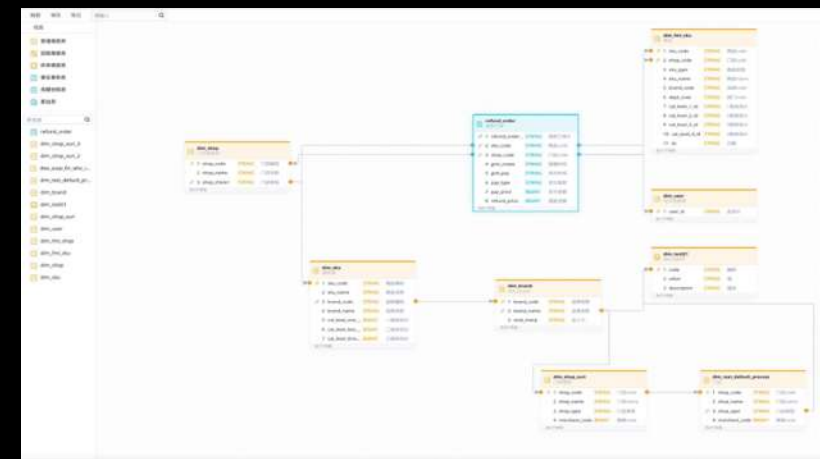
表血缘



SQL血缘



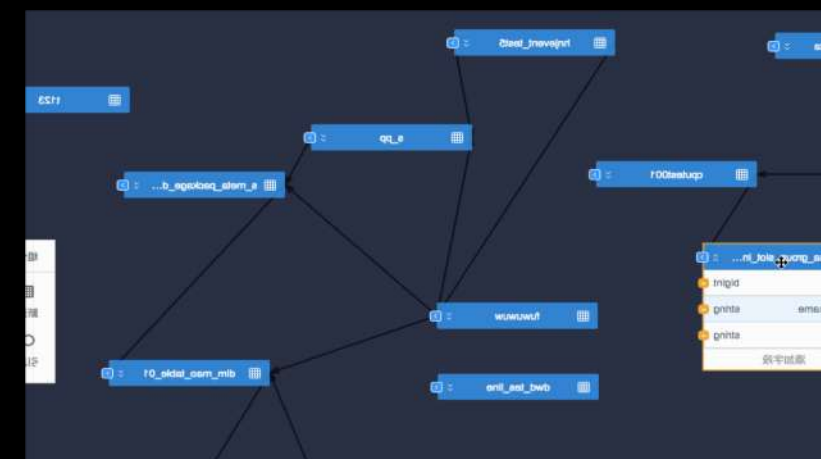
算子编排



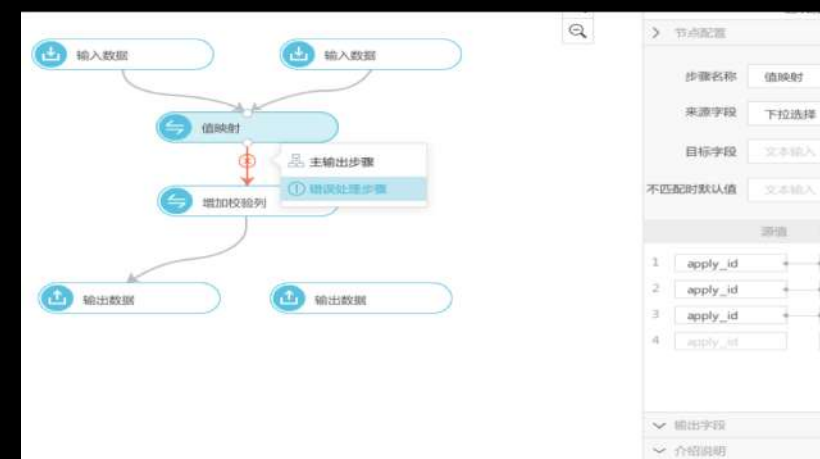
维度建模



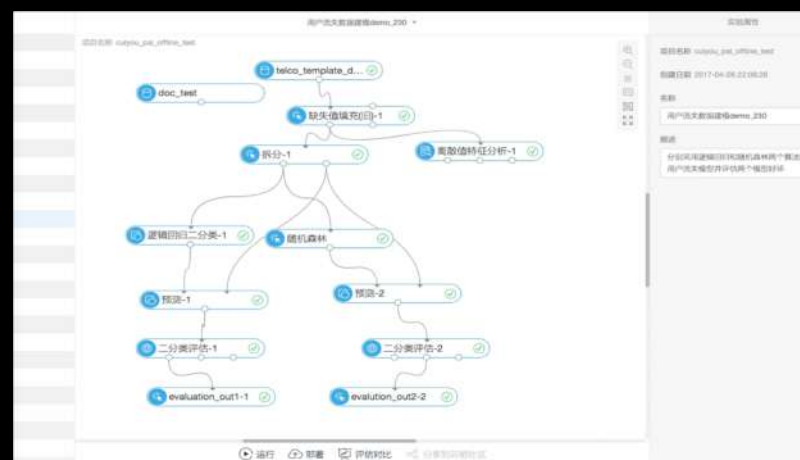
指标编排



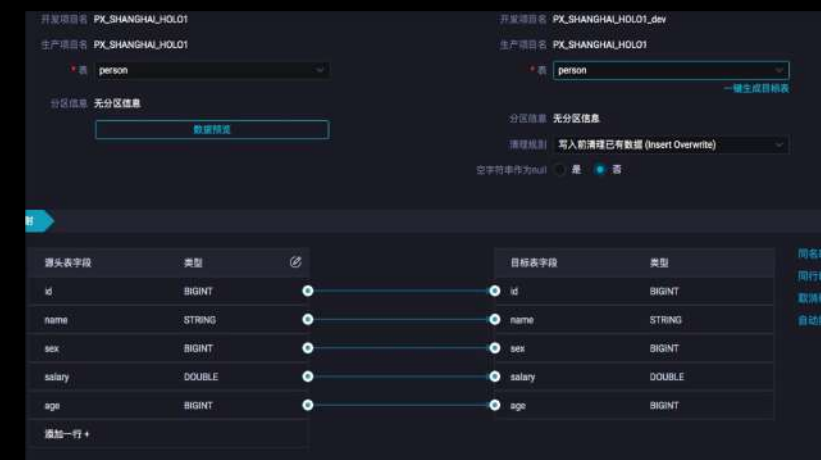
关系建模



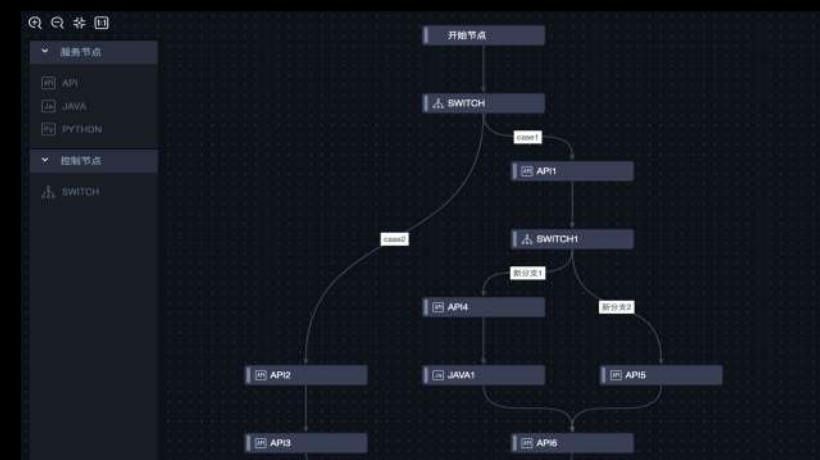
数据源编排



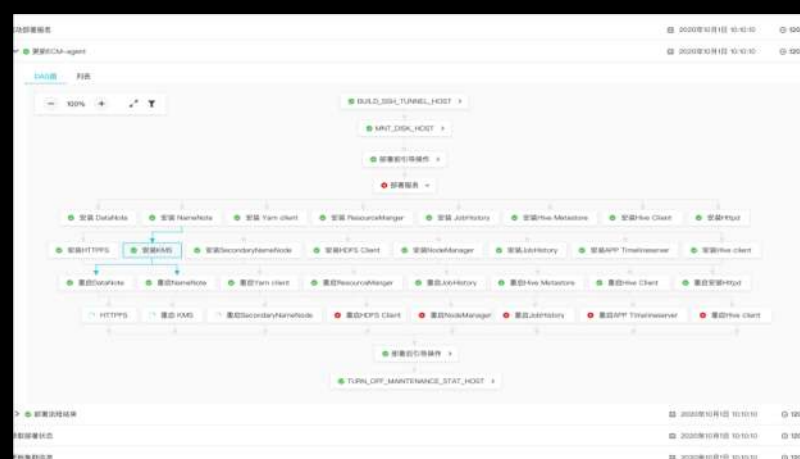
模型编排



字段映射



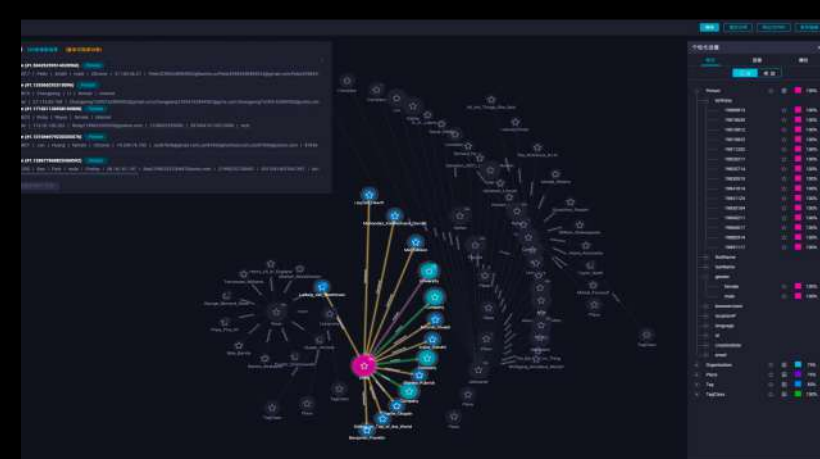
服务编排



集群调度



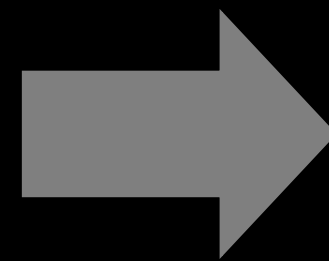
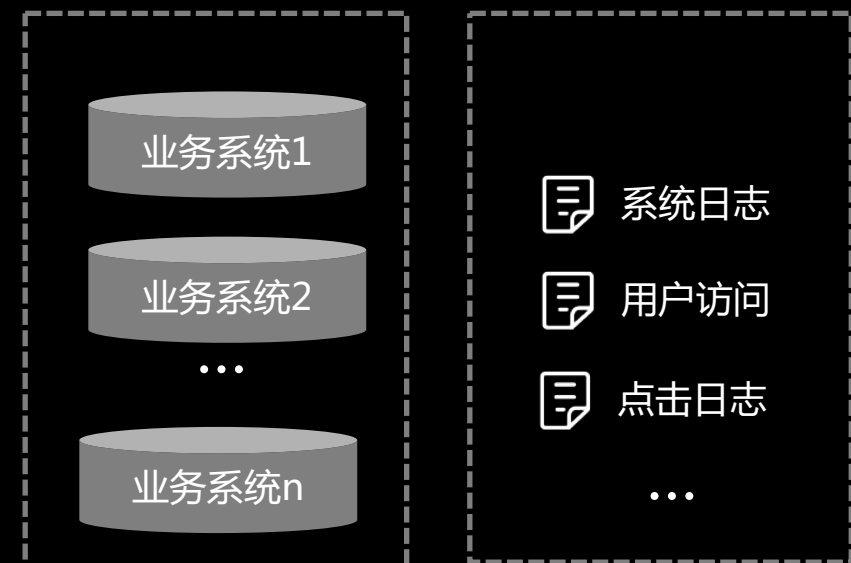
节点调度



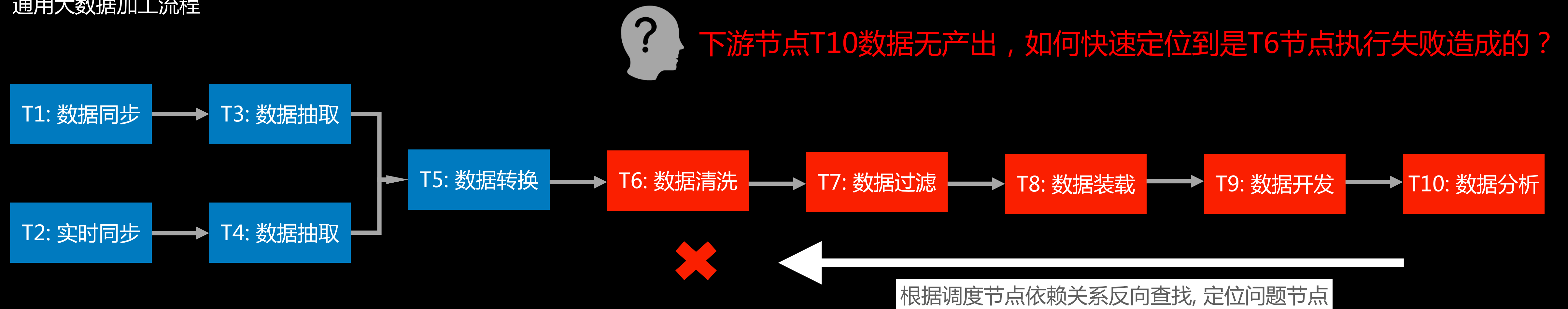
关系分析

- 阿里云飞天大数据平台 是阿里巴巴10年大数据建设最佳实践的结晶，每天有数万名数据和算法工程师正在使用飞天大数据平台，承载了阿里巴巴99%的数据业务构建。同时广泛应用于城市大脑、数字政府、电力、金融、新零售、智能制造、智慧农业等各个领域的大数据建设。
- DataWorks. 作为整个飞天大数据的重要组成部分，它基于MaxCompute/EMR/MC-Hologres等大数据计算引擎，为客户提供专业高效、安全可靠的一站式大数据开发与治理平台。

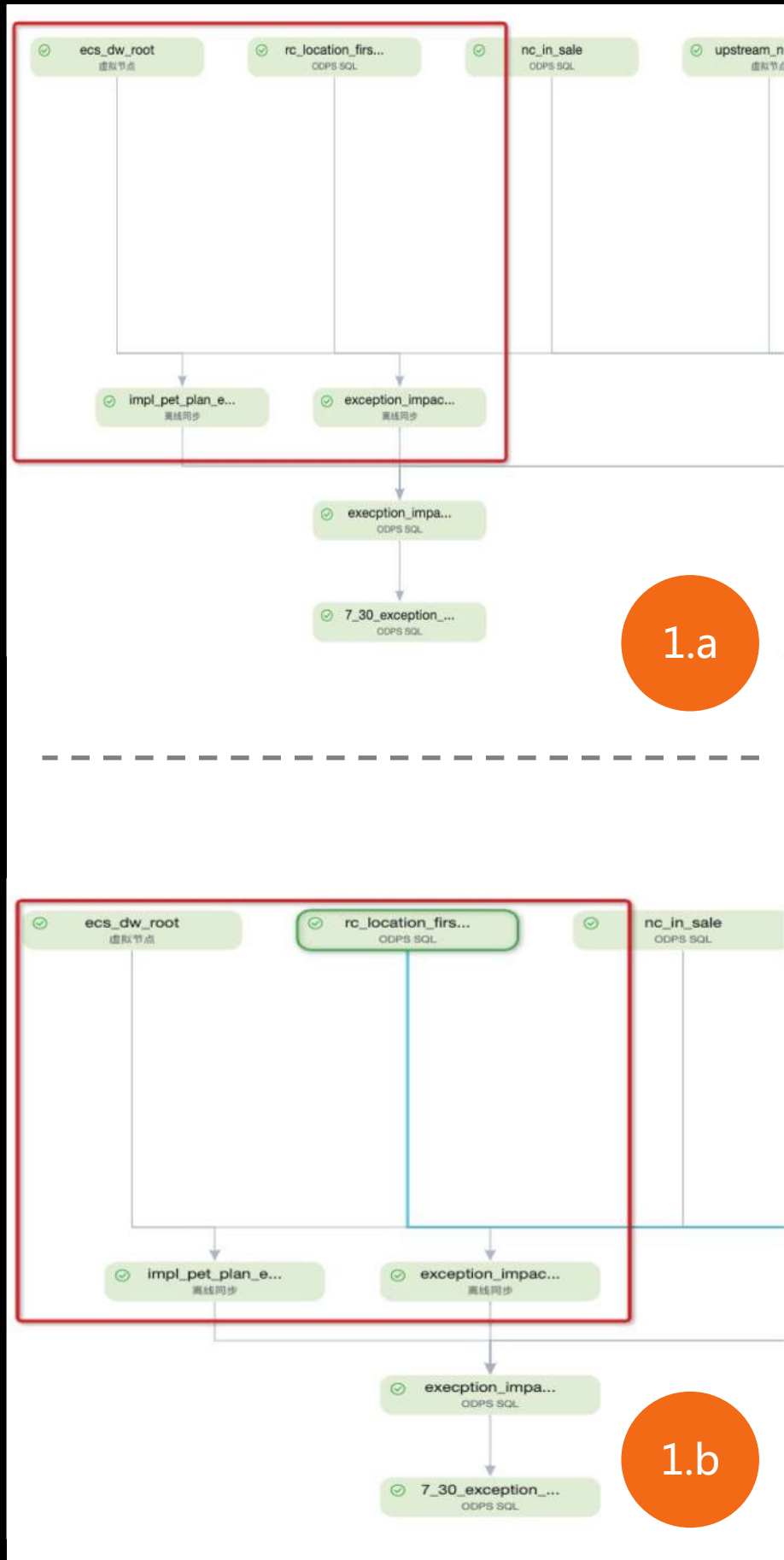
业务数据



通用大数据加工流程



使用DAGRE布局 + 视觉通道映射（颜色）技术绘制节点依赖关系



数据失真

1.a 画布节点依赖关系 1.b 真实数据无节点无任何依赖; 2. 下游节点过多，无法一屏展示; 3. 节点区分度不高、节点数过多、关系复杂溯源难度高

真实场景：使用DAGRE布局 + 视觉通道映射（颜色）技术绘制节点依赖关系



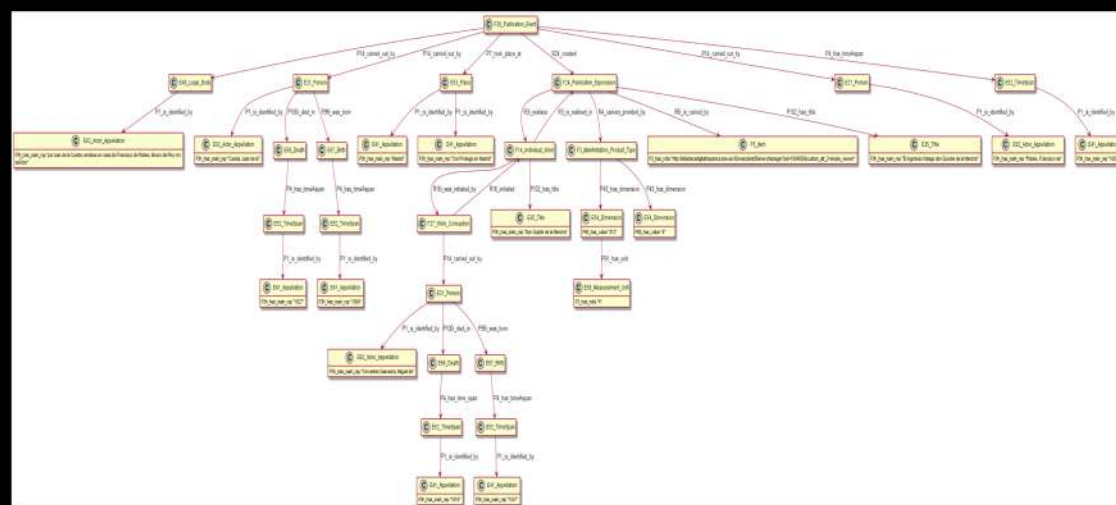
视觉通道映射\有限的信息密度\循环依赖.....

看不见的

- 数据庞大：节点数超过**千万**
- 同层宽度：同层节点数多，目前最多**十万+**
- 关系复杂：依赖复杂，跨周期、小时依赖等

真实场景：使用DAGRE布局 + 视觉通道映射（颜色）技术绘制节点依赖关系

离线渲染技术

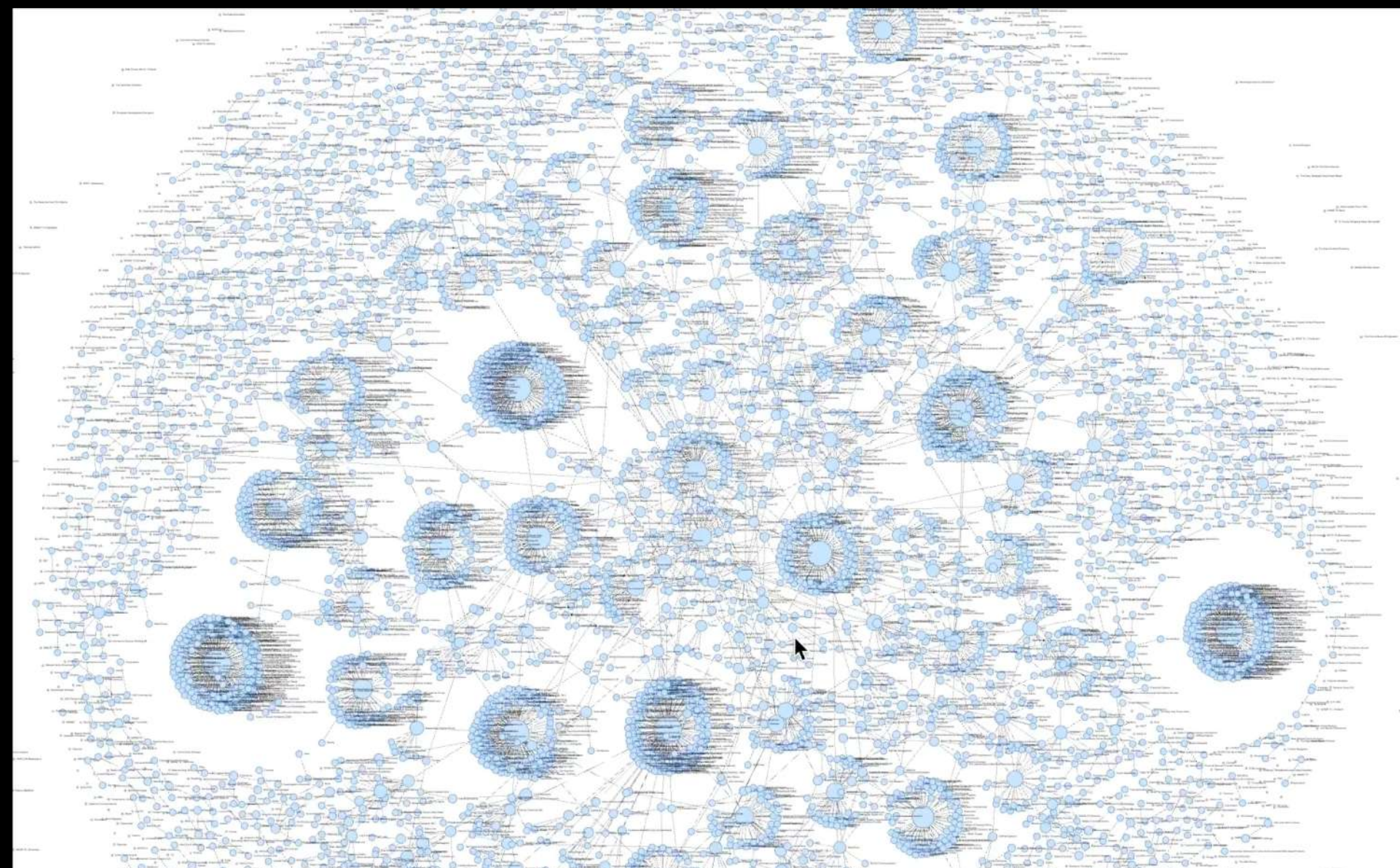


服务端渲染 (PlanUml)



Offline Canvas

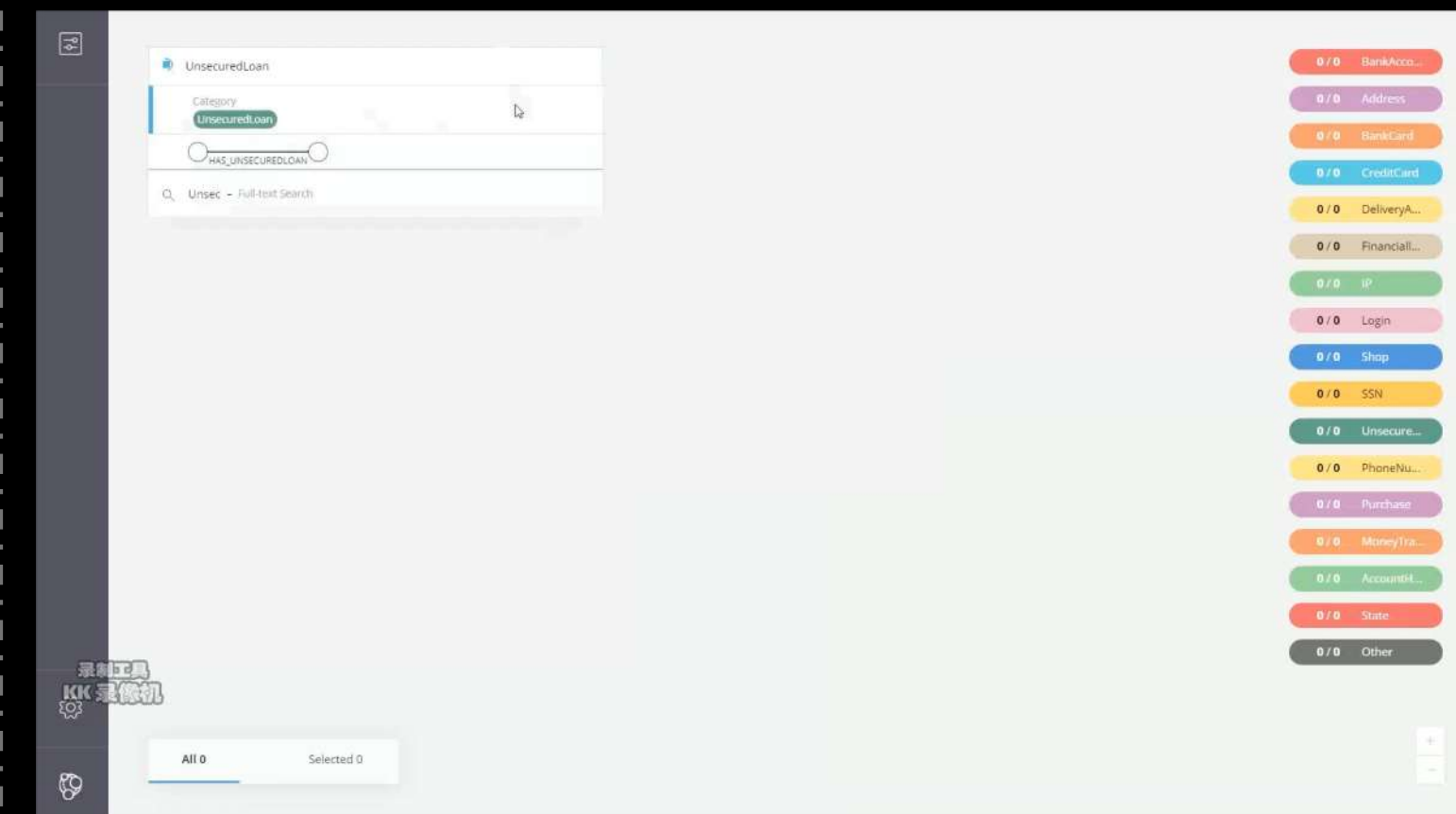
局部渲染技术



追求极致性能的@antv/G6引擎：

1. 支持局部渲染技术，只渲染可视窗口内信息
2. 使用脏矩阵渲染，在每一帧中进行最小区域的重绘
3. 使用 R-tree 空间索引加速区域查询

双引擎切换能力

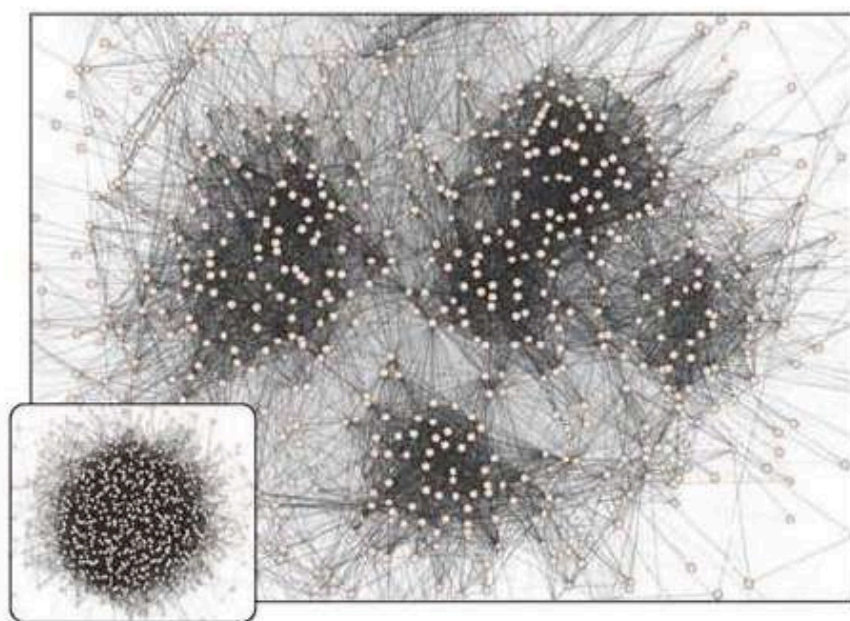


Neo4J Bloom的双引擎机制：

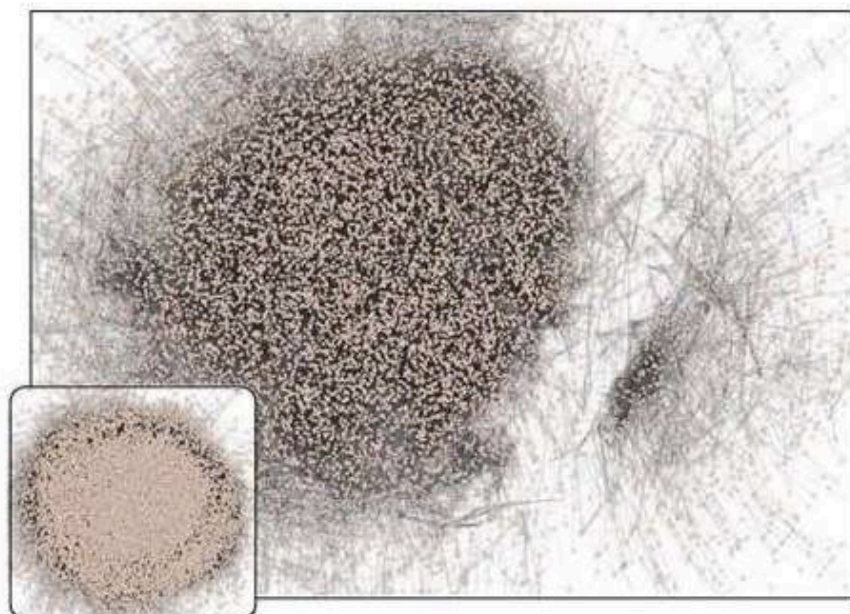
- 全局视角采用GPU Powered Rendering能力，展示更多数据
- 局部视角采用Canvas Rendering能力，展现更丰富的细节



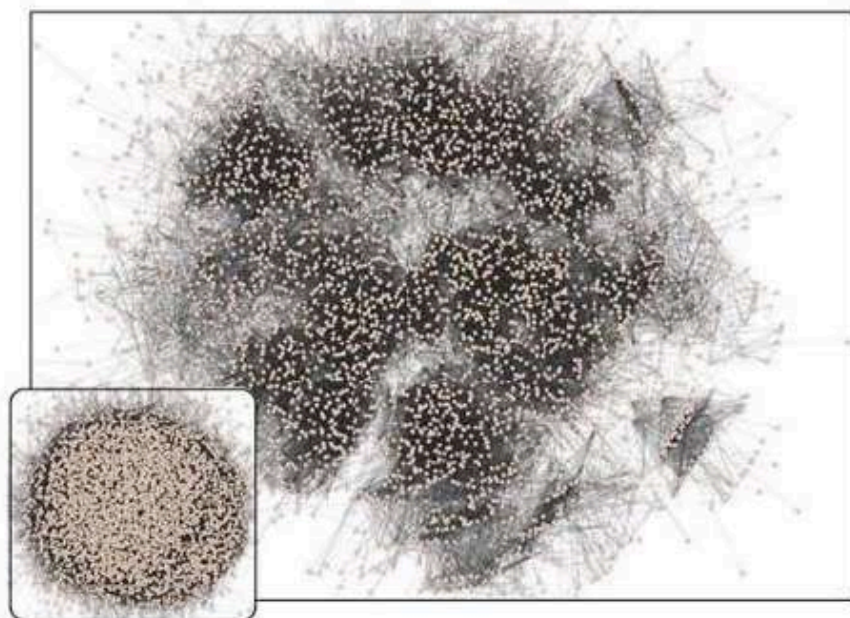
图简化技术



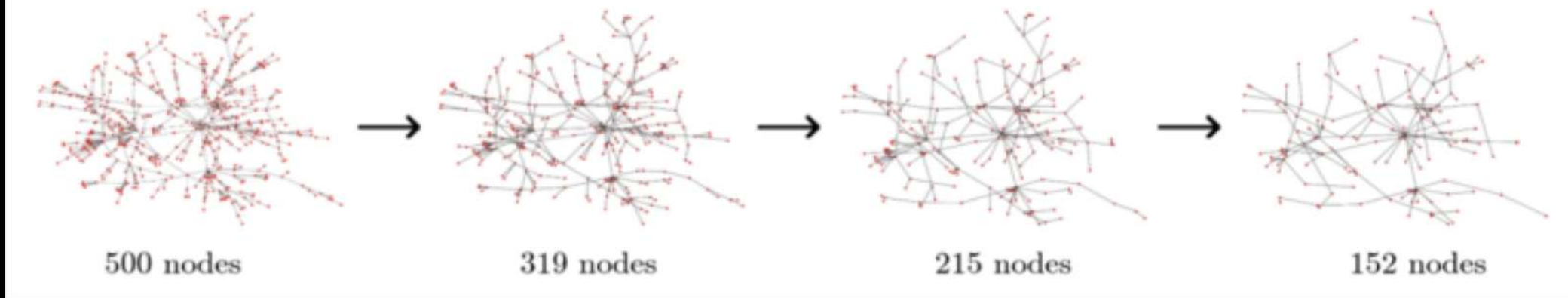
(a) Caltech36, $|V| = 762$, $|E| = 16k$



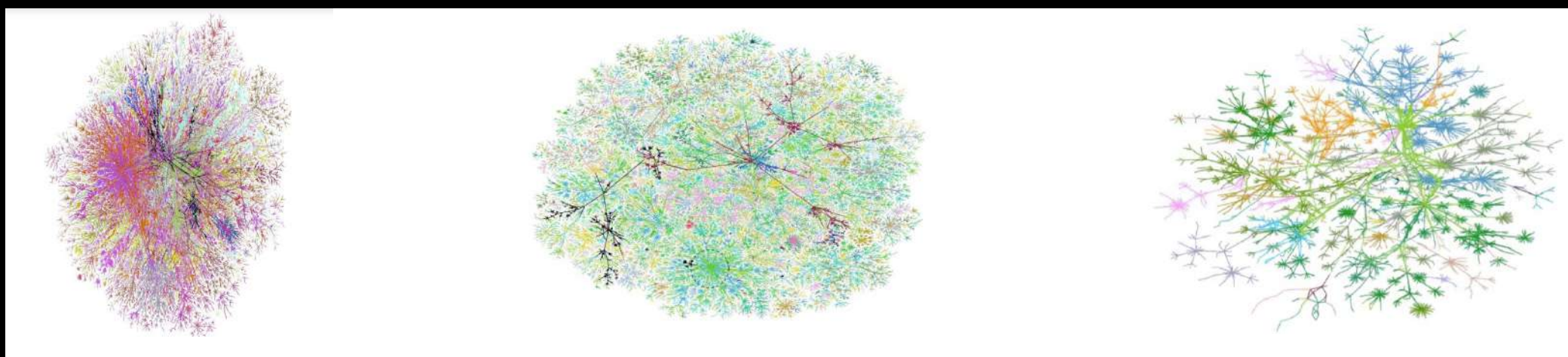
(b) Harvard1, $|V| = 15k$, $|E| = 800k$



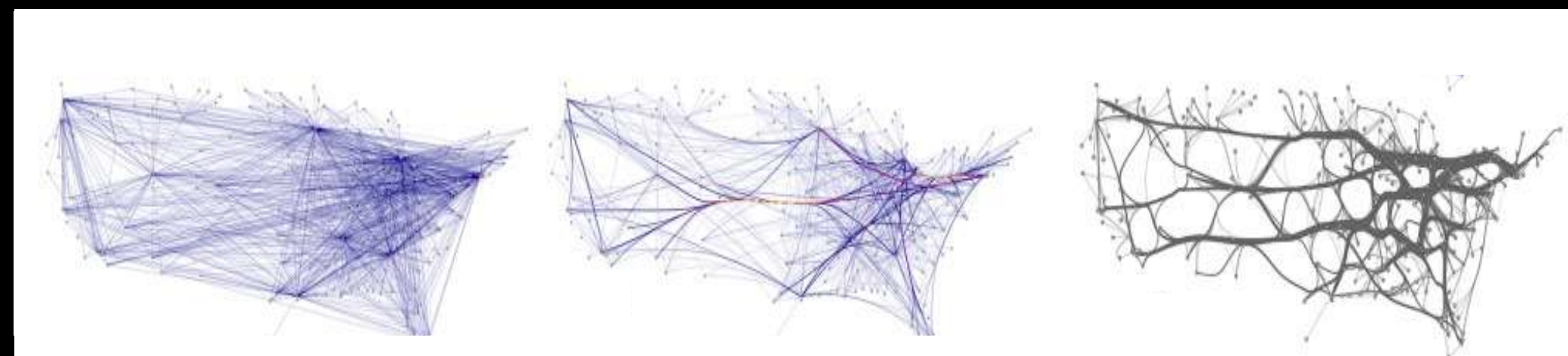
(c) Rice31, $|V| = 4k$, $|E| = 97k$



✓ 点聚合 (Node Aggregation) 的目的是提取出大规模数据不同“粒度”的图，前端首先展示最粗粒度的图，终端用户通过放大、下钻的方式查看更细粒度的图/子图



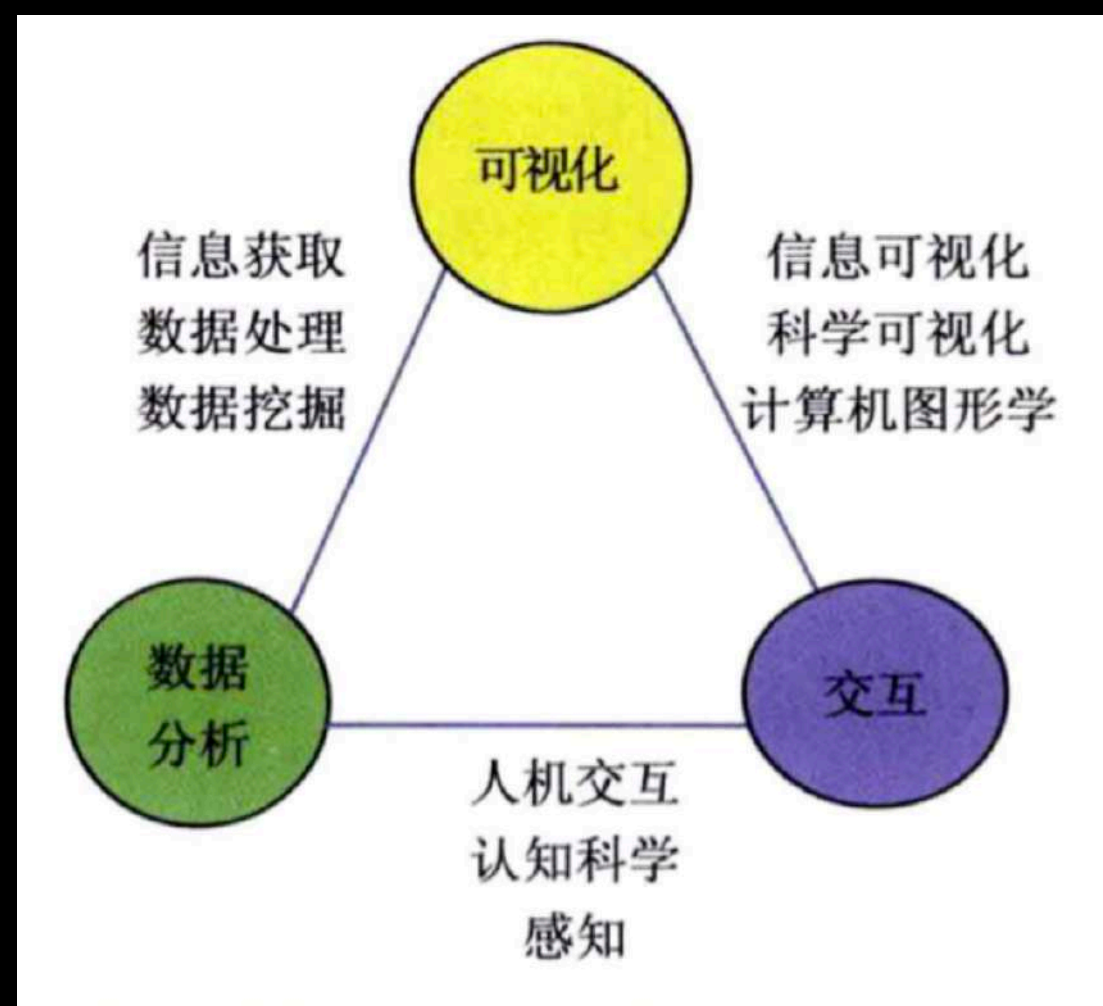
边剪枝 (edge bundling) 通过反向删除算法、Prim算法构建最小生成树，通过只保留关键路径减少视觉混乱



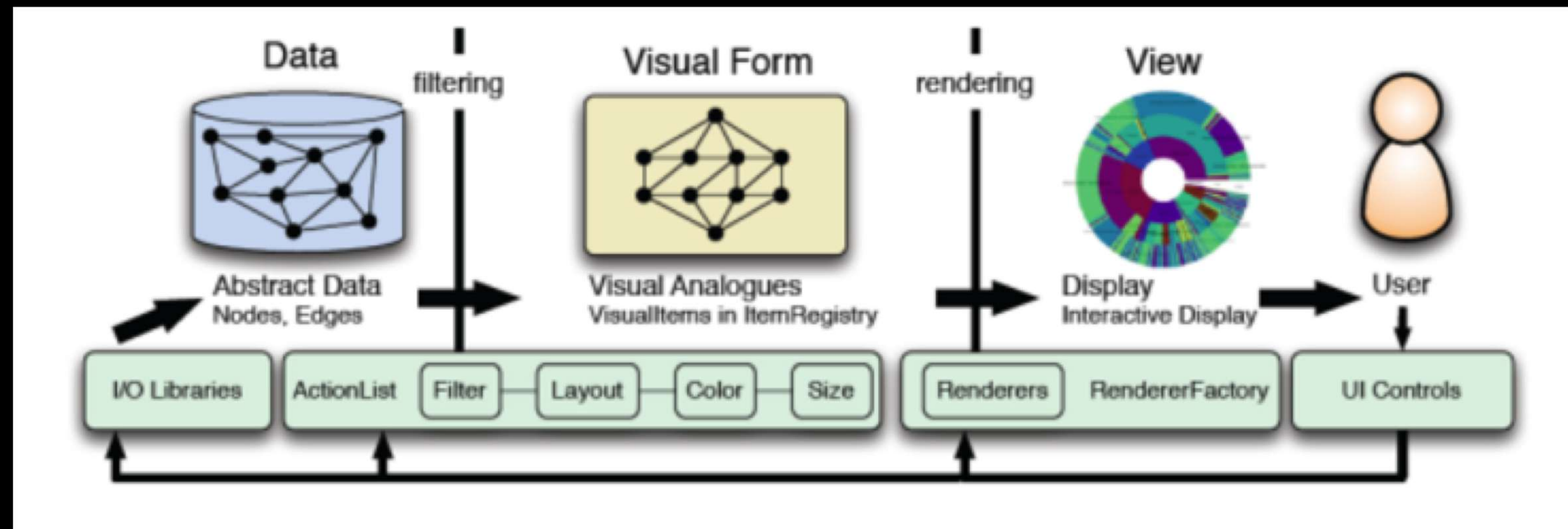
边绑定 (Edge Bundling) 边缘捆绑技术被设计为在视觉上将相似的边缘捆绑在一起，以减少图中的视觉杂乱。这种捆绑可以帮助突出重要的边缘模式，并且常常使人更容易发现有趣的连接或重要的数据

“毛团”现象导致关系网络可读性下降

过渡追求渲染能力，导致的“毛团”现象



可视分析组成

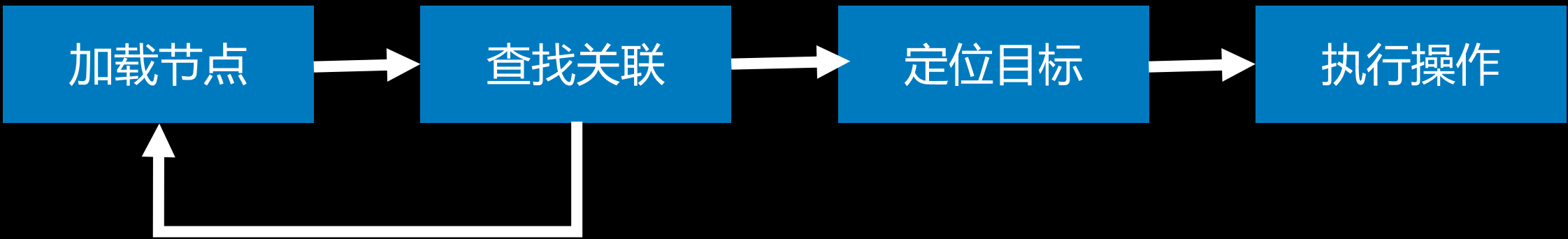
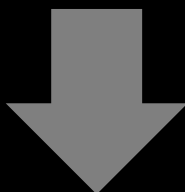


数据可视化流程

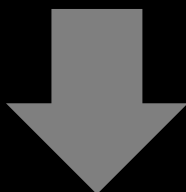
- 传统的可视化：大多立足于先验知识，解决一些具体且可预期的任务时有一定的优势。
- 可视分析：将「人」这个独特因素融入进了数据分析的流程中。通过人机交互技术，将更多的重点放在了人的意会与推理上，让人在分析任务中参与了主要的分析与决策过程。有了可视分析的支持，我们就能够在很多单纯依赖算法分析无法解决问题的场景中，进行可视知识发现，获取有价值的信息。

- 查看节点的上下游依赖
- 定位上游的阻塞节点
- 查看上游节点的产出
- 判断节点对下游的影响
- 查看业务团队的节点影响范围
- 暂停节点运行
- 重新运行节点
- 问题节点重新执行
- ...

用户诉求

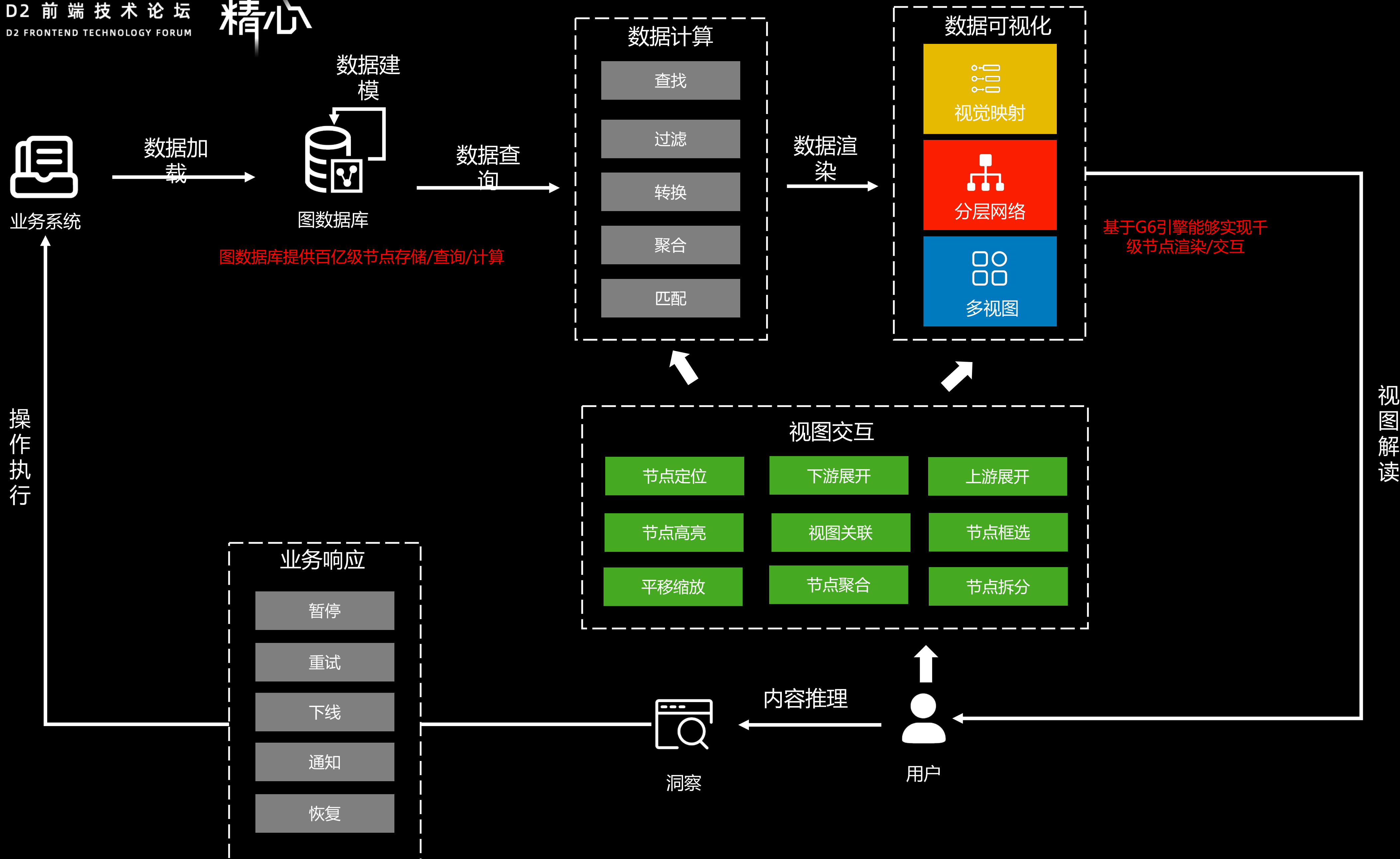


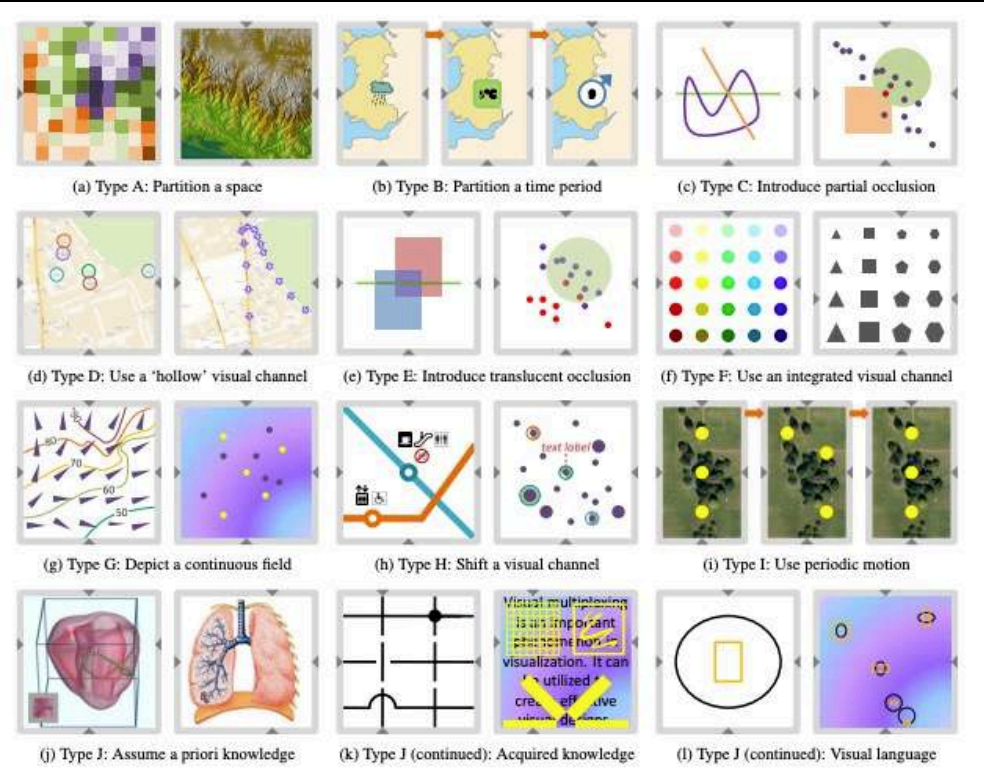
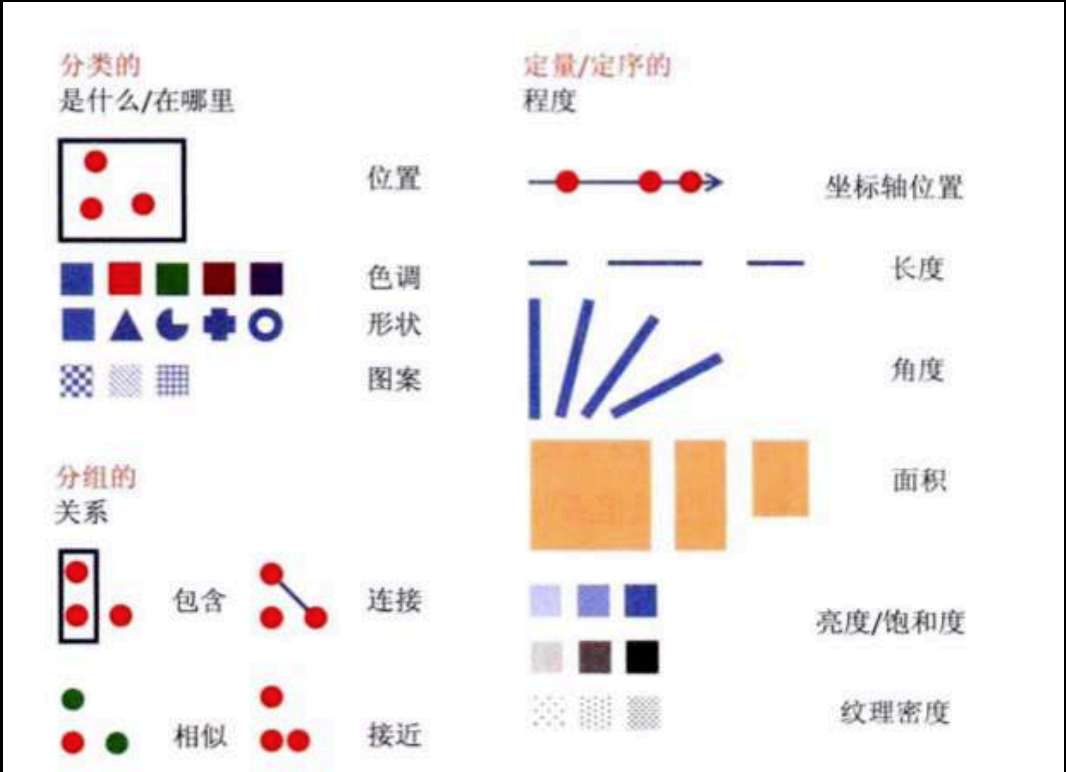
交互流程



视觉映射 节点布局 节点定位 节点聚合 节点过滤 视角切换 视图关联 节点框选

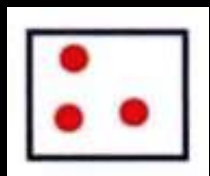
可视分析技术





有限的视觉通道，以及表现力排序

视觉多通技术实现多可视化信息堆叠



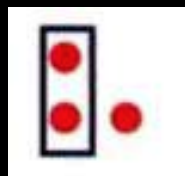
位置



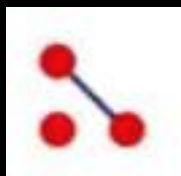
色调



亮度/饱和度



包含

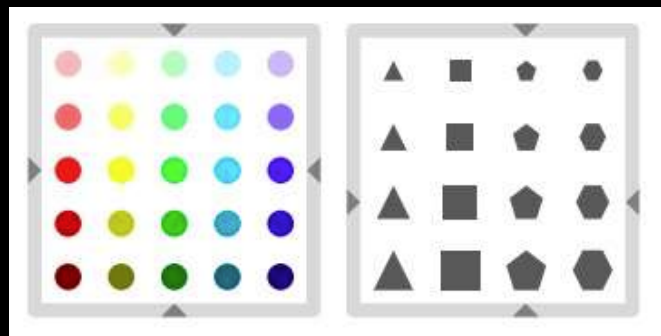


关联

+



空间分割



多通道整合

可视分析－视觉映射

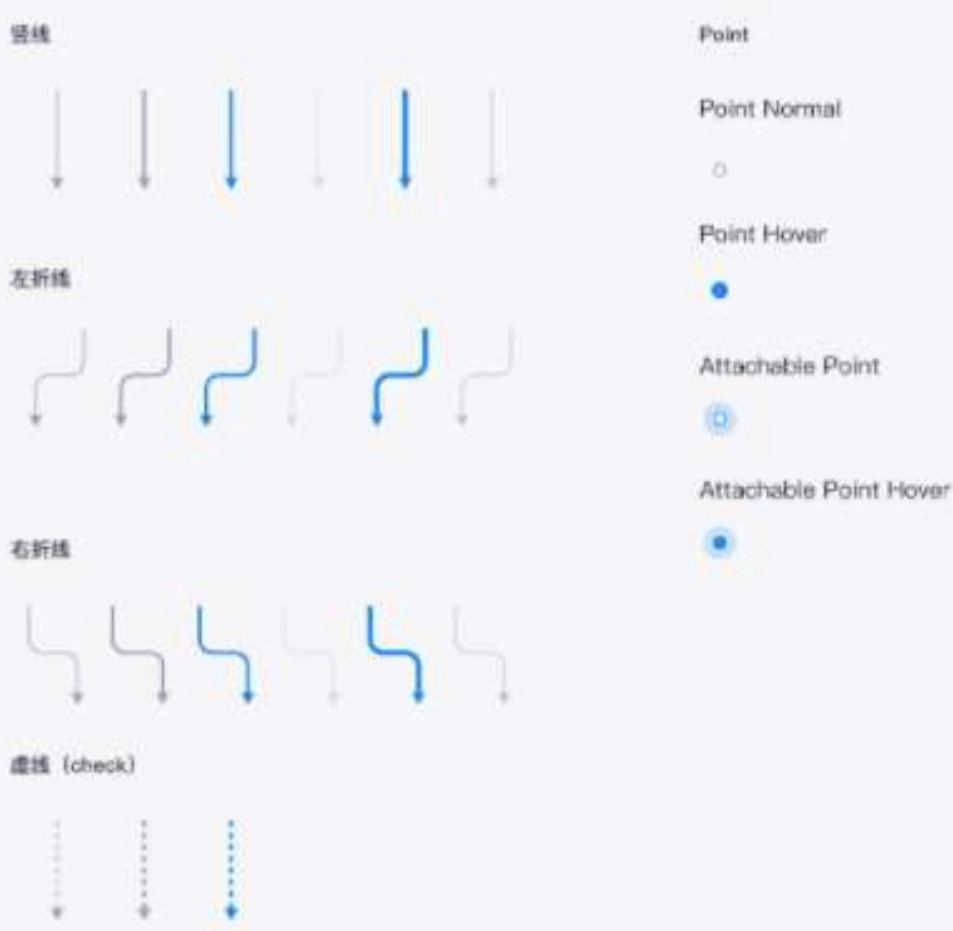
任务节点样式



任务组合节点样式



边样式

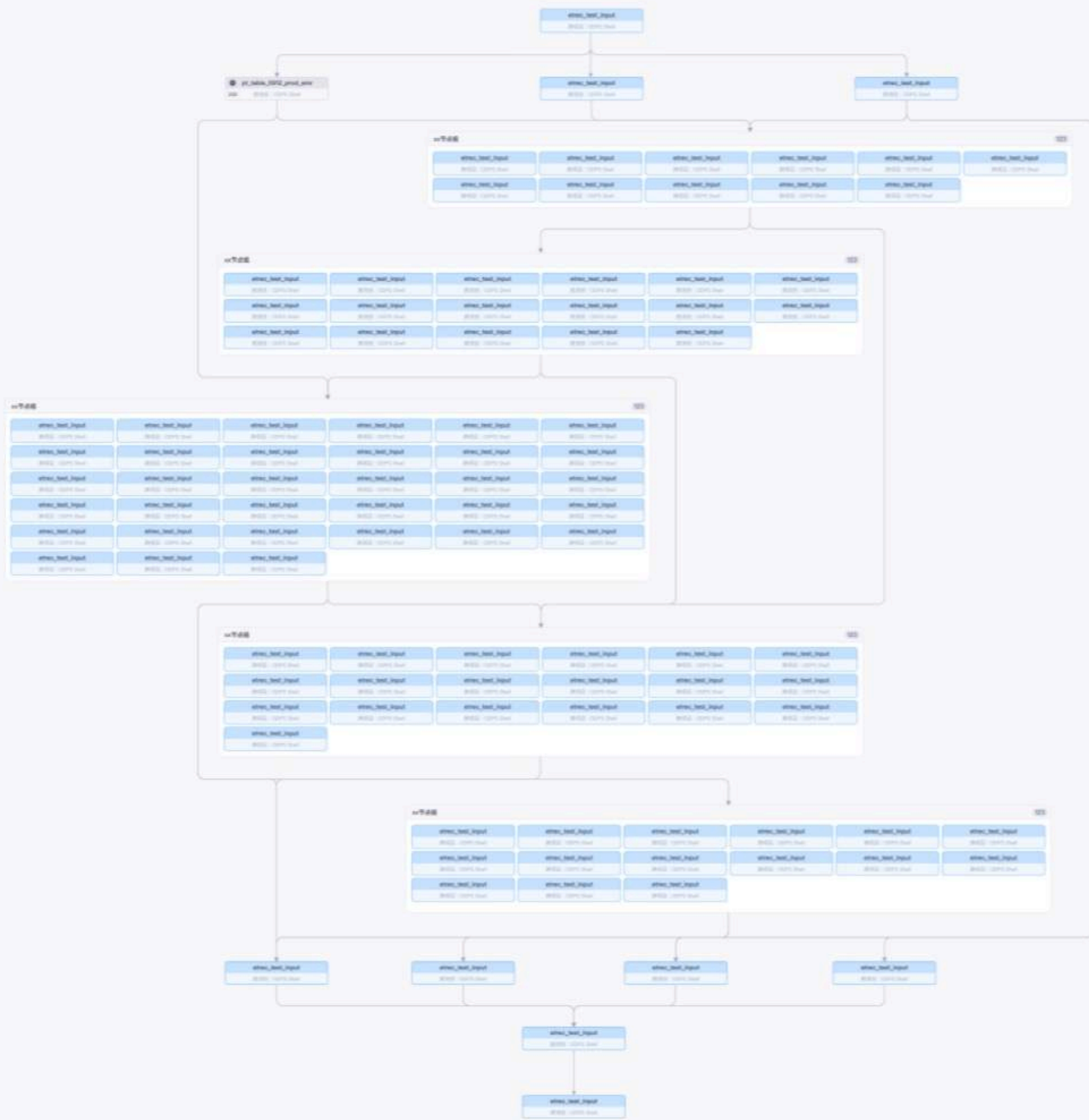


右键



组合样式



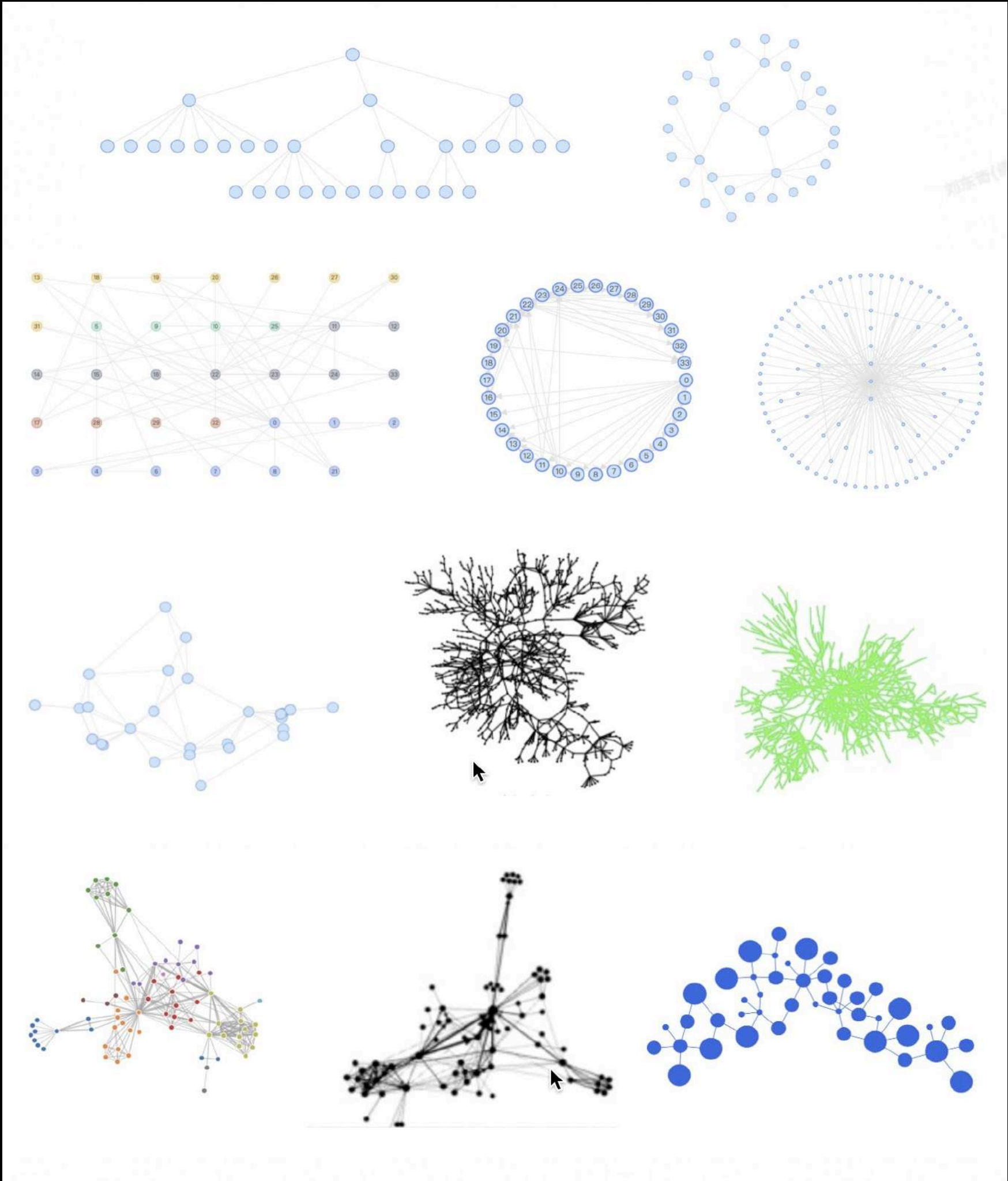


分层布局

策略型

方程优化

力导布局



图可视化常见布局策略

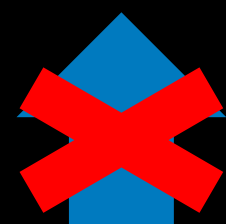
业务场景	分析目的	技术方案
套现网络/闭环网络	看清楚是否存在环	•圆形布局 circle •力导布局 force
分层网络/流程网络	看清楚层次	•有向分层 dagre •树形力导 force-tree
无序网络数据	看清楚 点和边，不要点线交叉	•力导布局 force •fruchterman
暴涨网络 / 坍塌网络	关注中心位置的情况	•同心圆布局 cencentric •径向布局 radial
团伙聚集网络	关系聚集，分类现象	•同心圆布局 cencentric •聚类布局 node combo •Circle packing
多边有向网络	看清楚 边的方向	•弦图布局 arc

常见布局策略应用场景



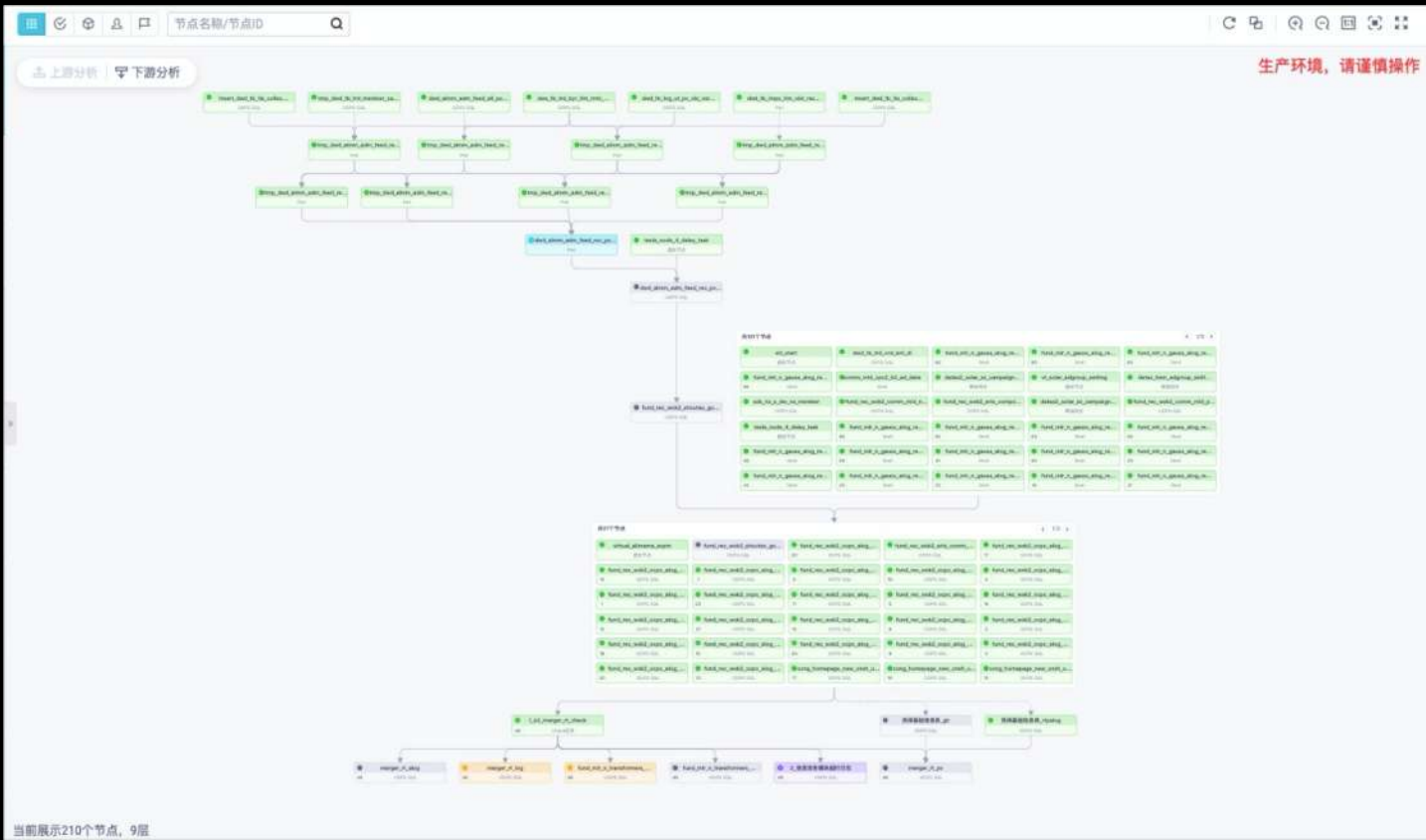
布局核心

- 清晰展示节点信息
- 清晰表示依赖关系
- 帮助用户快速分析问题

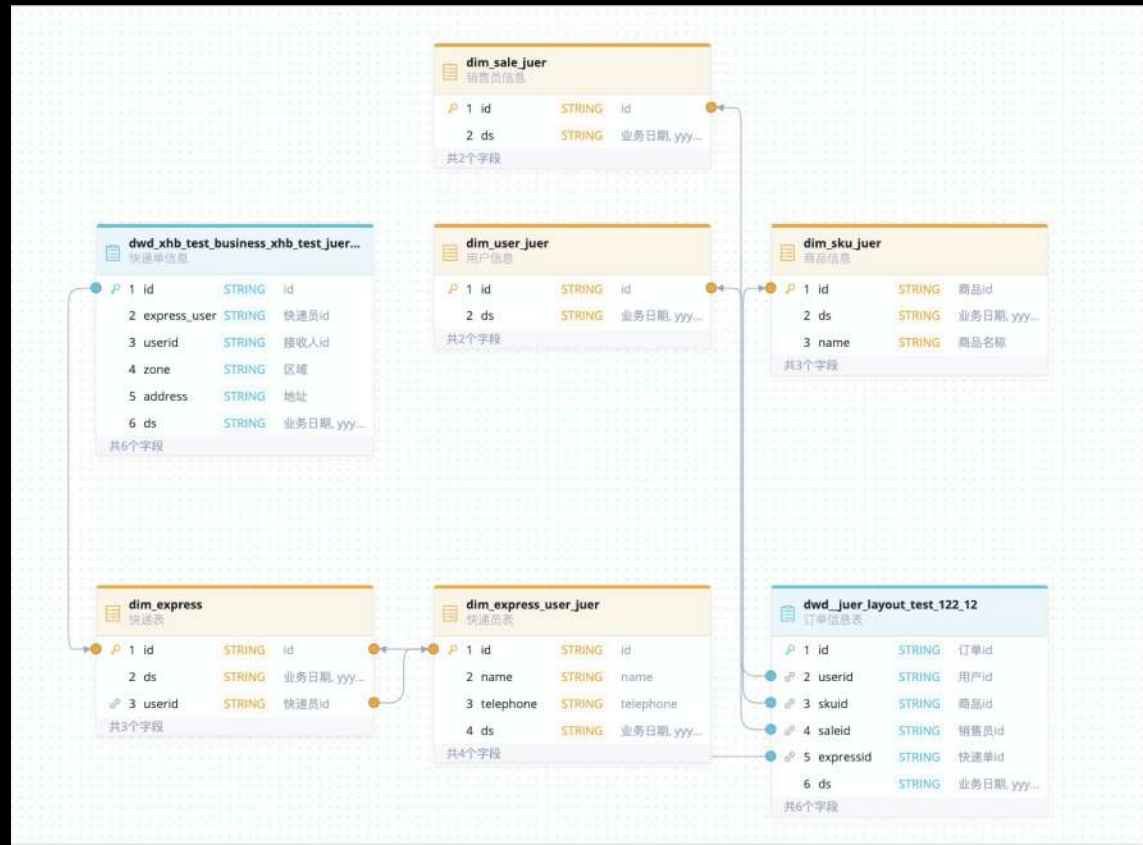


基于Sankey布局构建的分层网络

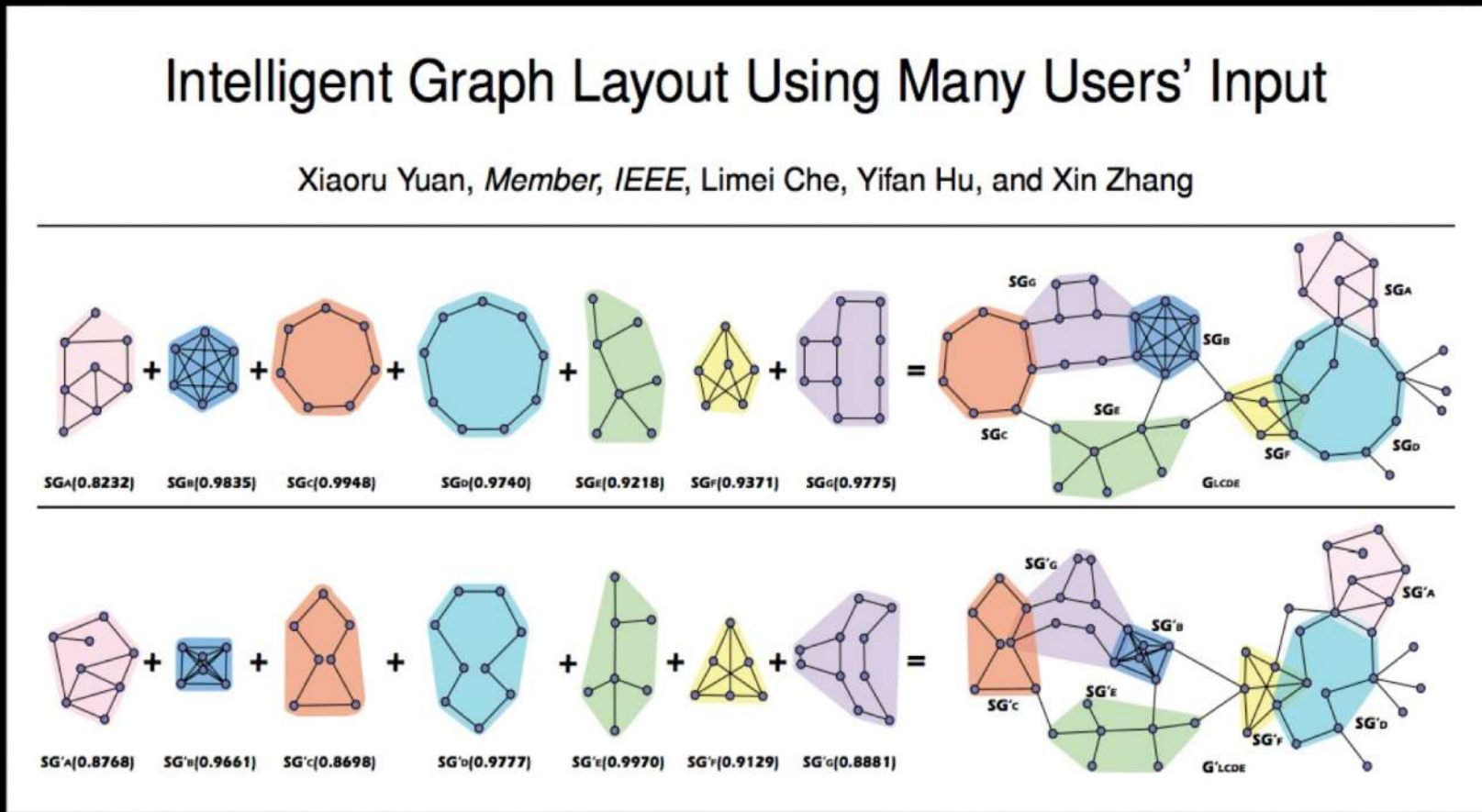
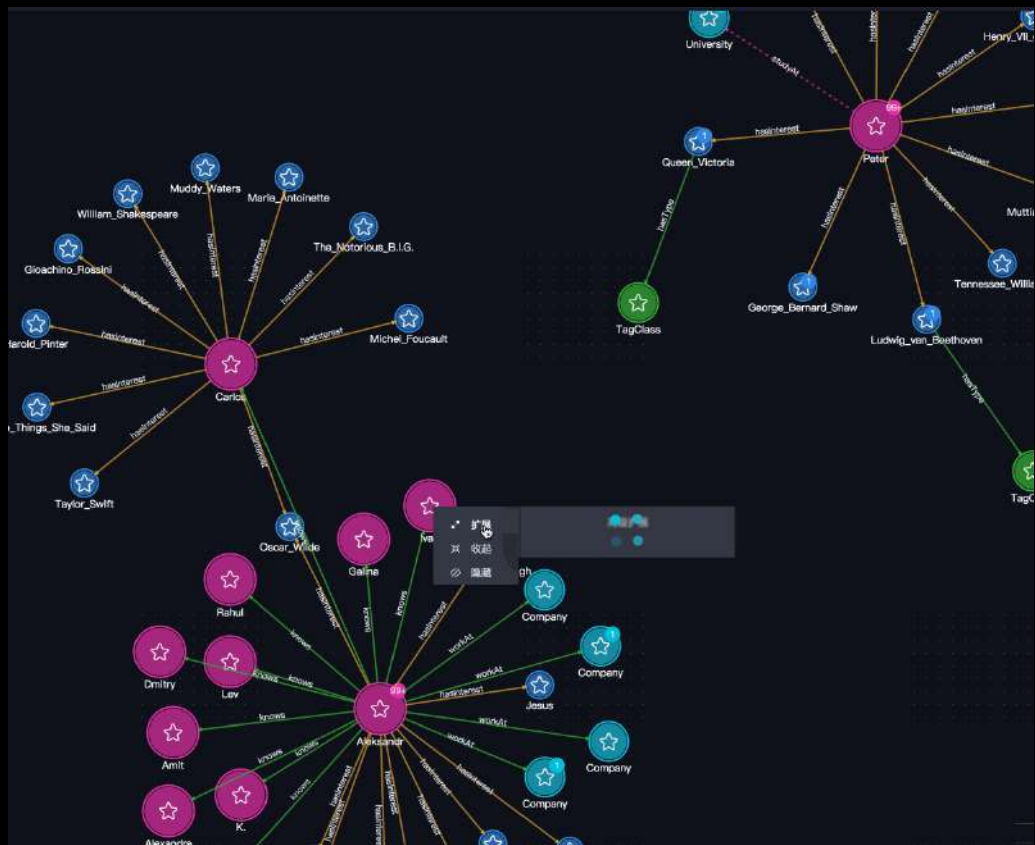
- 图宽特征下造成布局浪费
- Sankey布局缺陷，布局中不考虑边交叉问题。
- 线的路由算法，合并线后丢失方向信息。



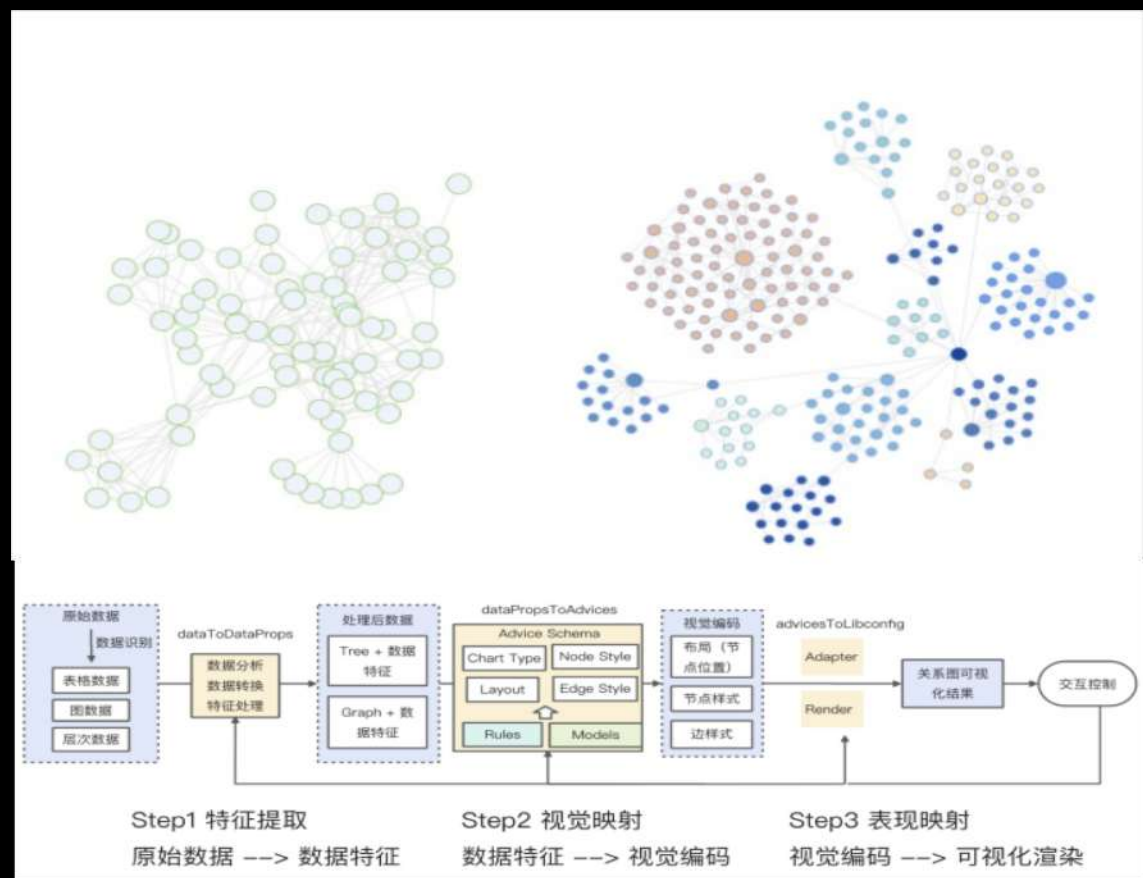
子图嵌入
DAGRE + GRID



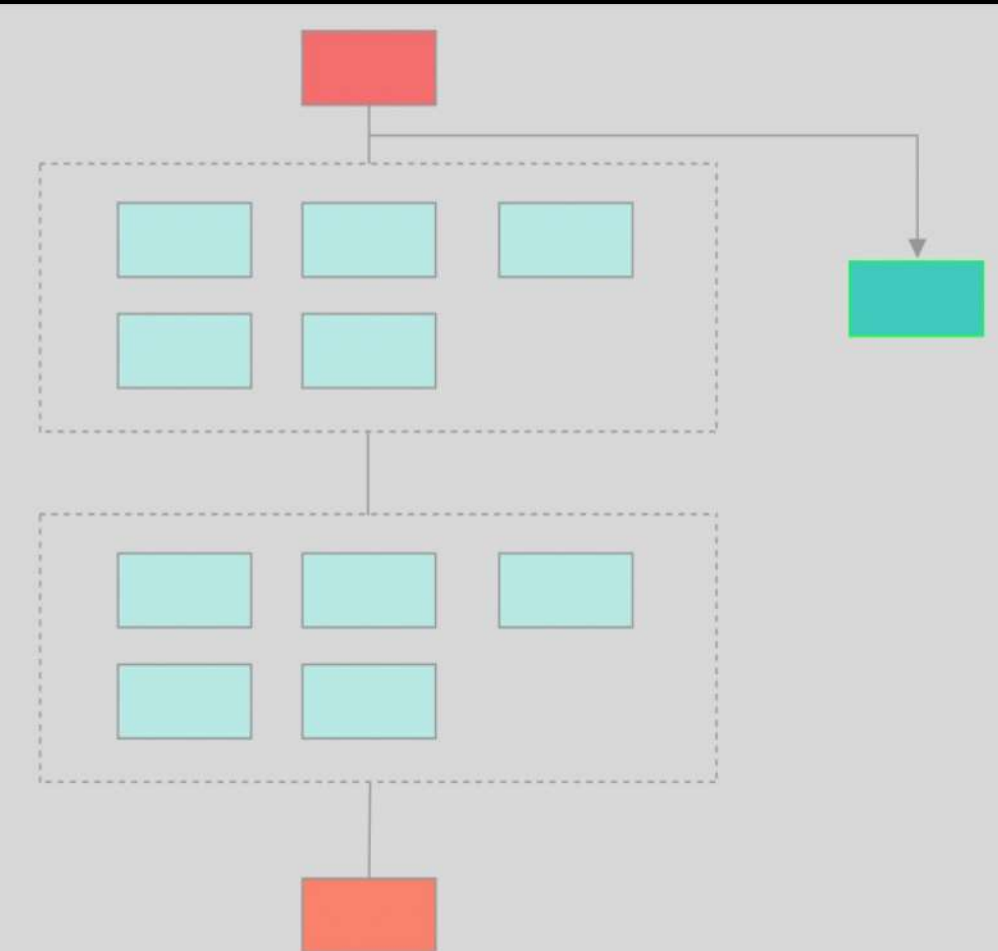
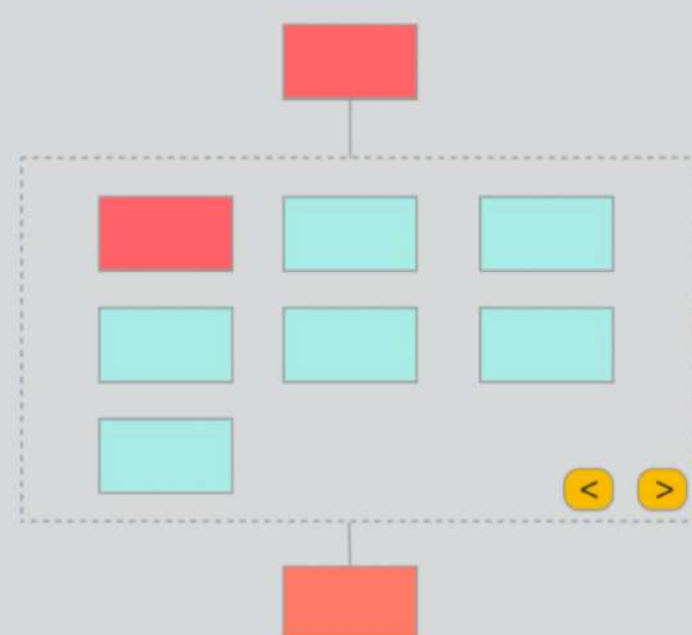
布局叠加



子图融合



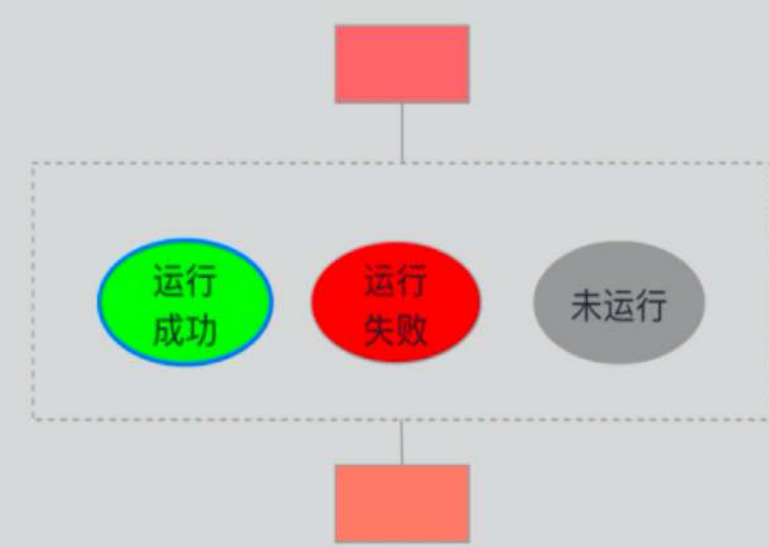
智能布局



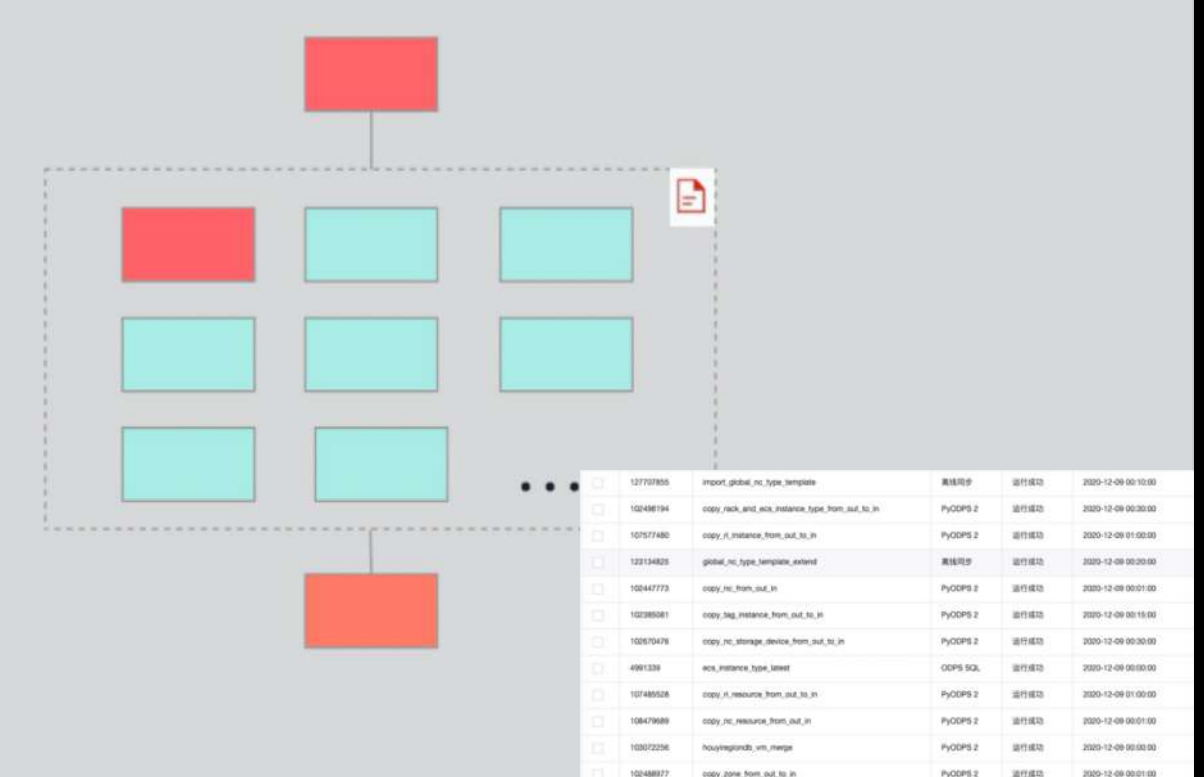
分组设计

- 同层节点置于一个分组中，支持换行翻页。
- 提供聚合工具，帮助用户快速分析。
- 通过辅助视图，展示全量节点。

聚合



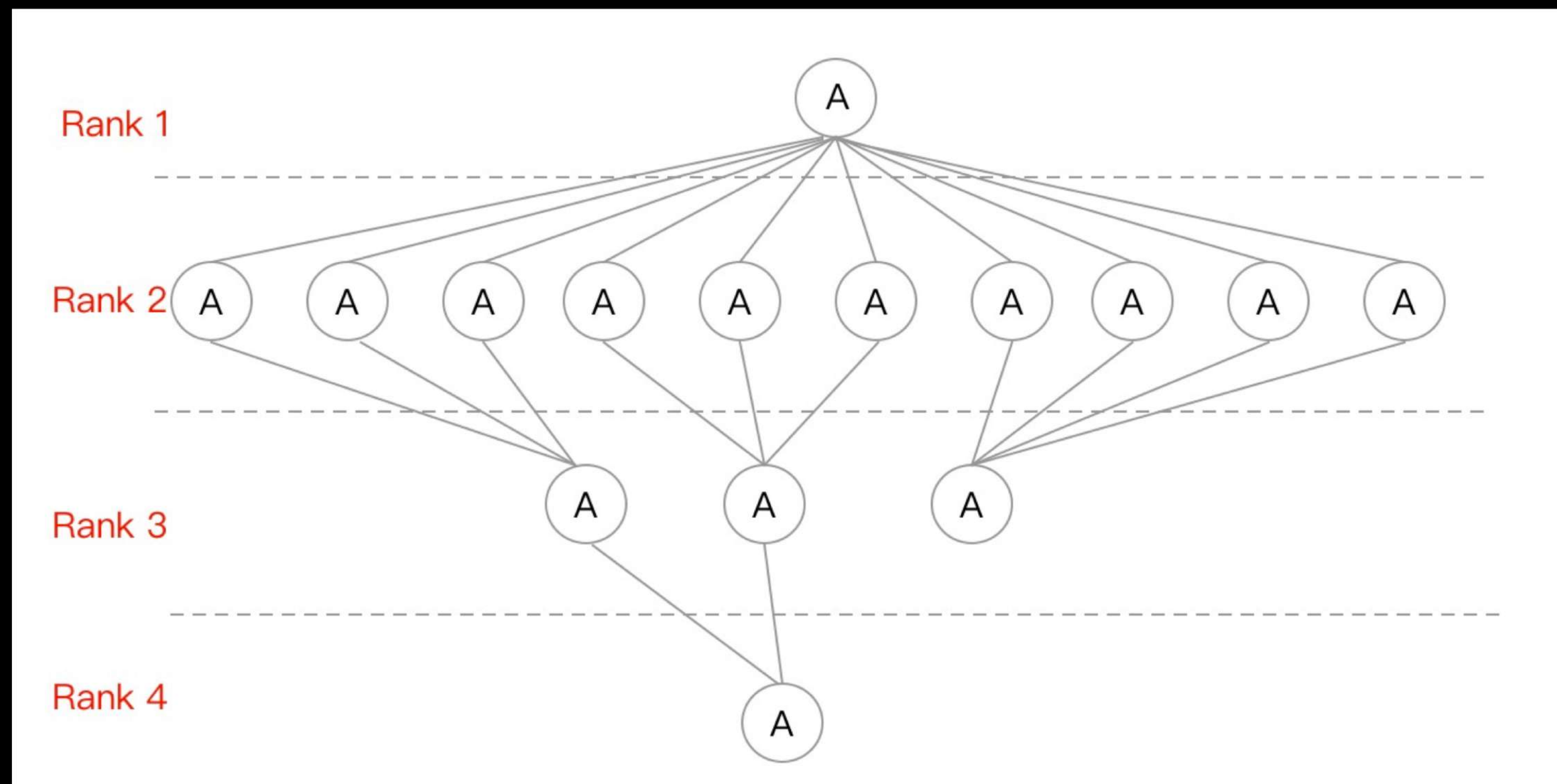
辅助视图



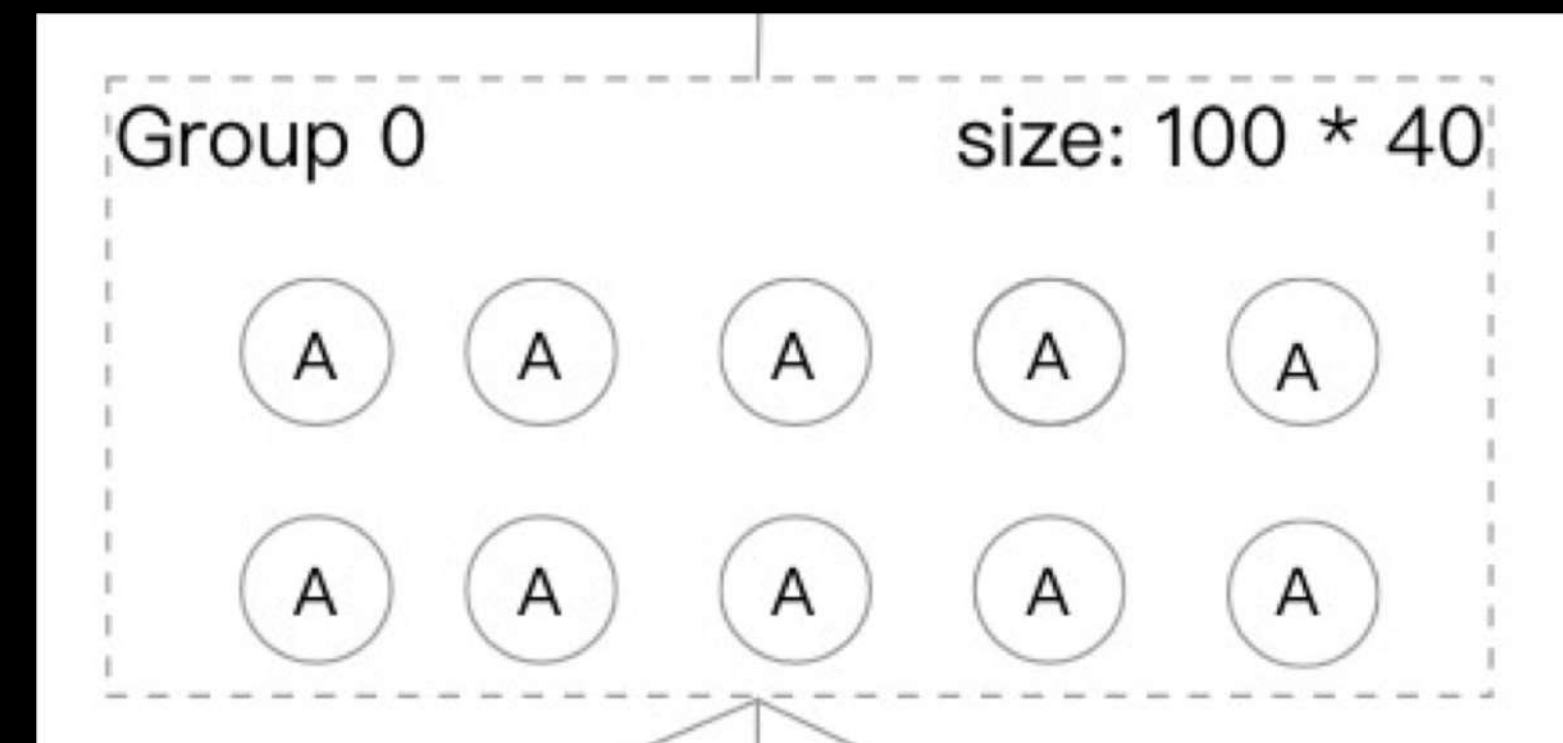
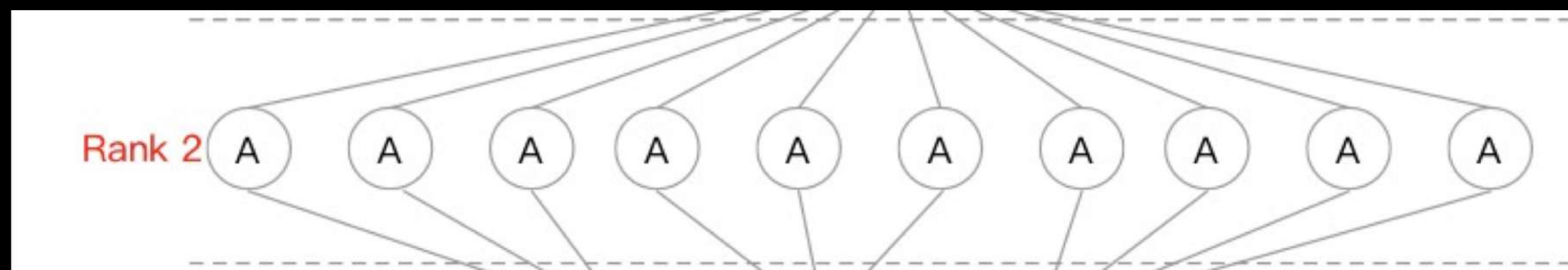
分层网络 + 分组设计

1. 节点分层：根据依赖关系进行分层，采用紧致树分层算法。
2. 虚拟分组：判断每层节点数量是否大于MAX_LEVEL，是则生成虚拟分组节点。
3. 节点布局：虚拟节点以及正常节点，进行Dagre布局。
4. 分组内嵌：根据虚拟节点的坐标，转换层内节点的相对坐标。

节点分层



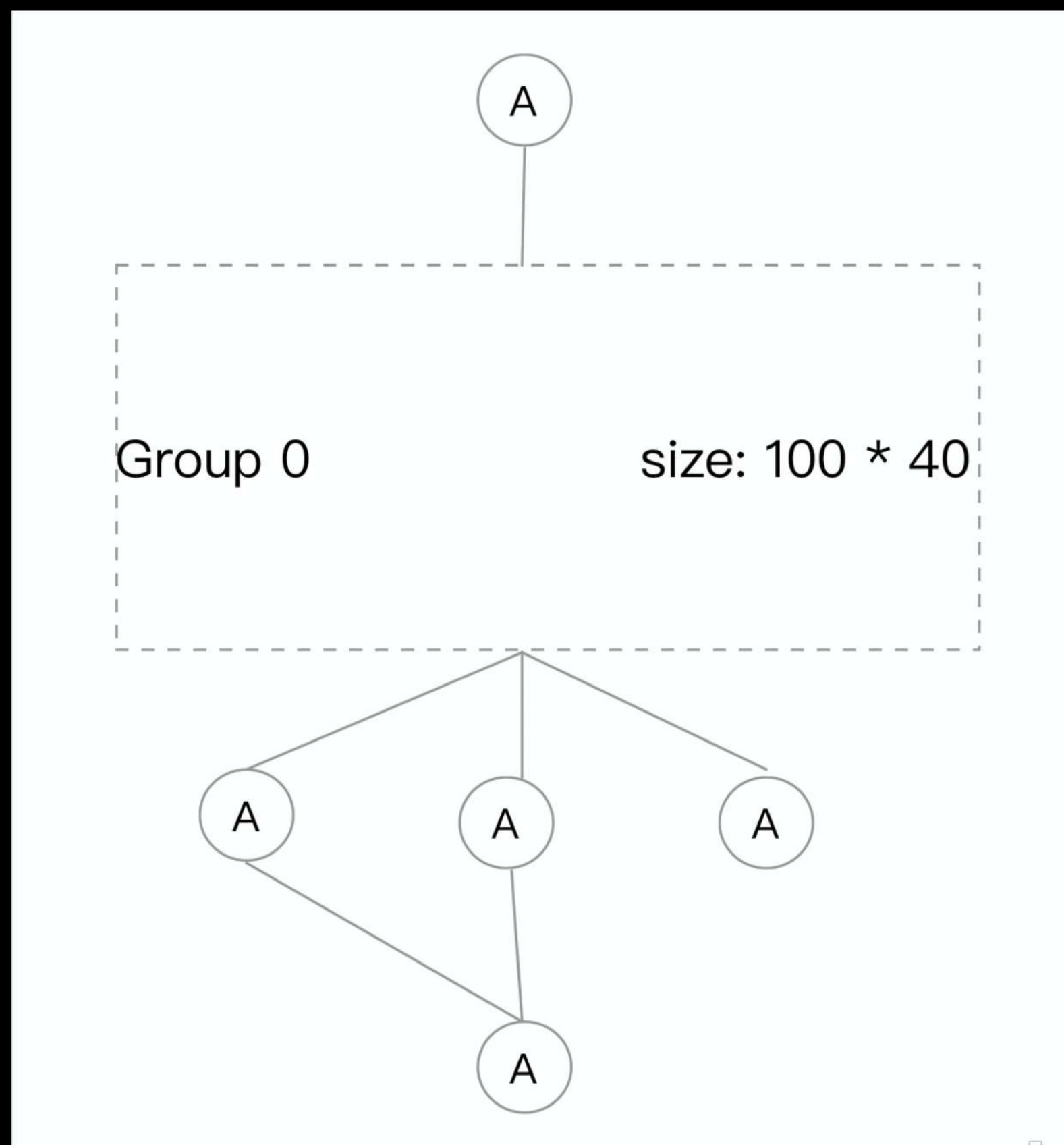
虚拟分组



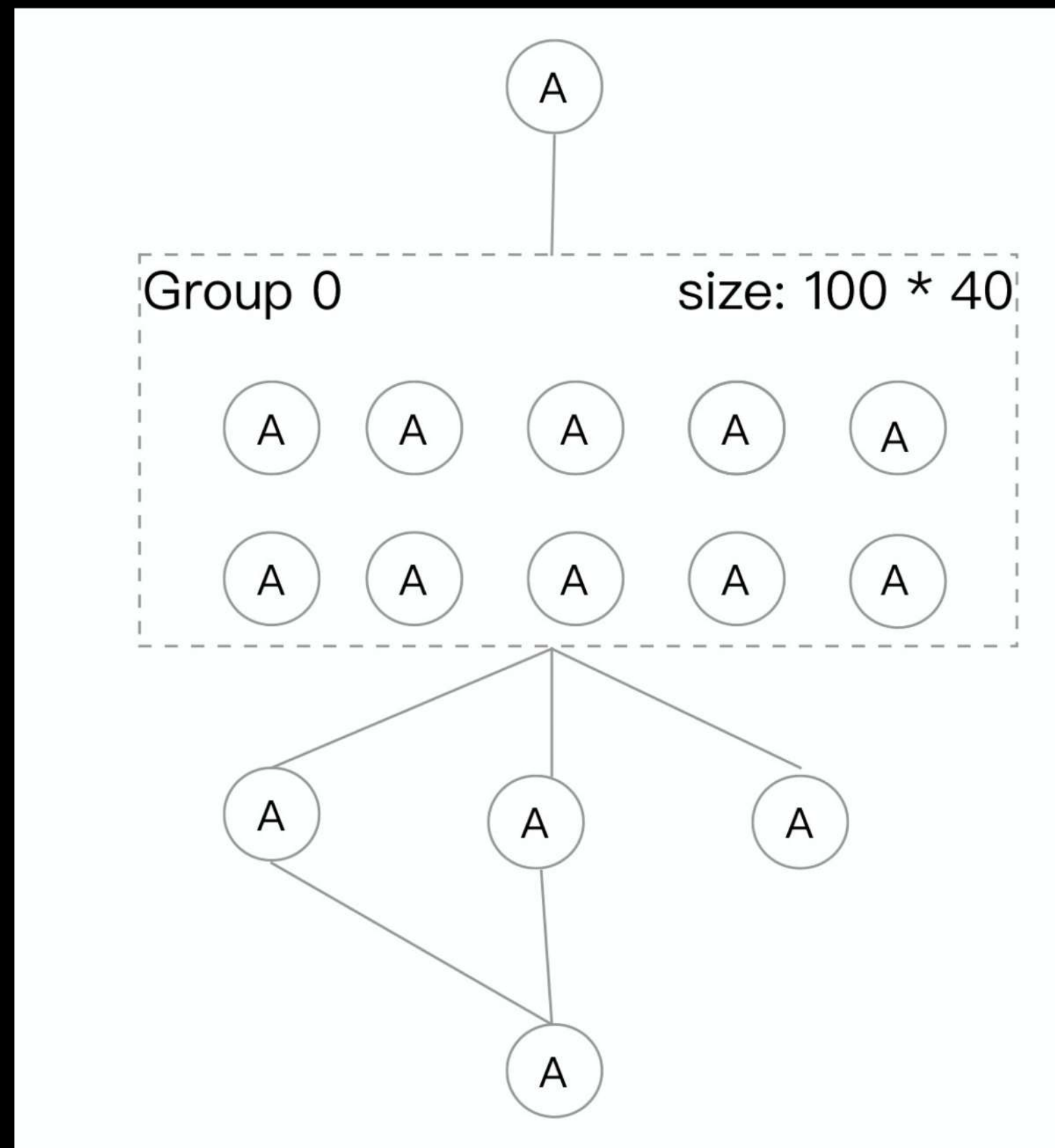
$$Width = (NodeWidth + NodeGap) \times Column + 2 \times GroupPadding$$

$$Height = (NodeHeight + NodeGap) \times (NodeNum \div Column) + 2 \times GroupPadding$$

Dagre布局



节点内嵌



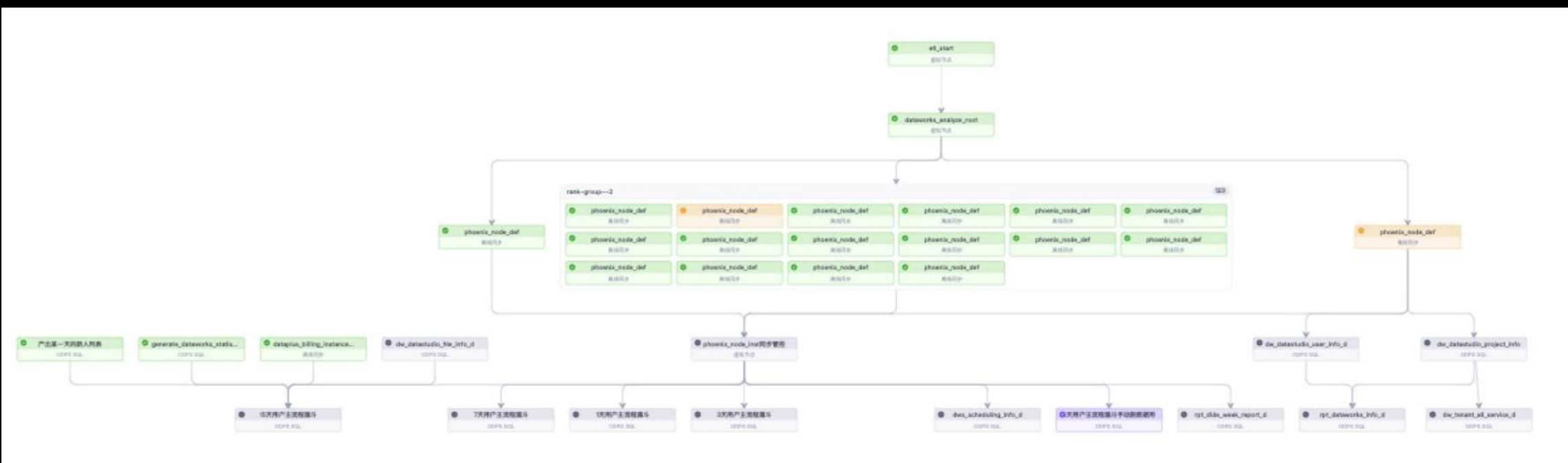
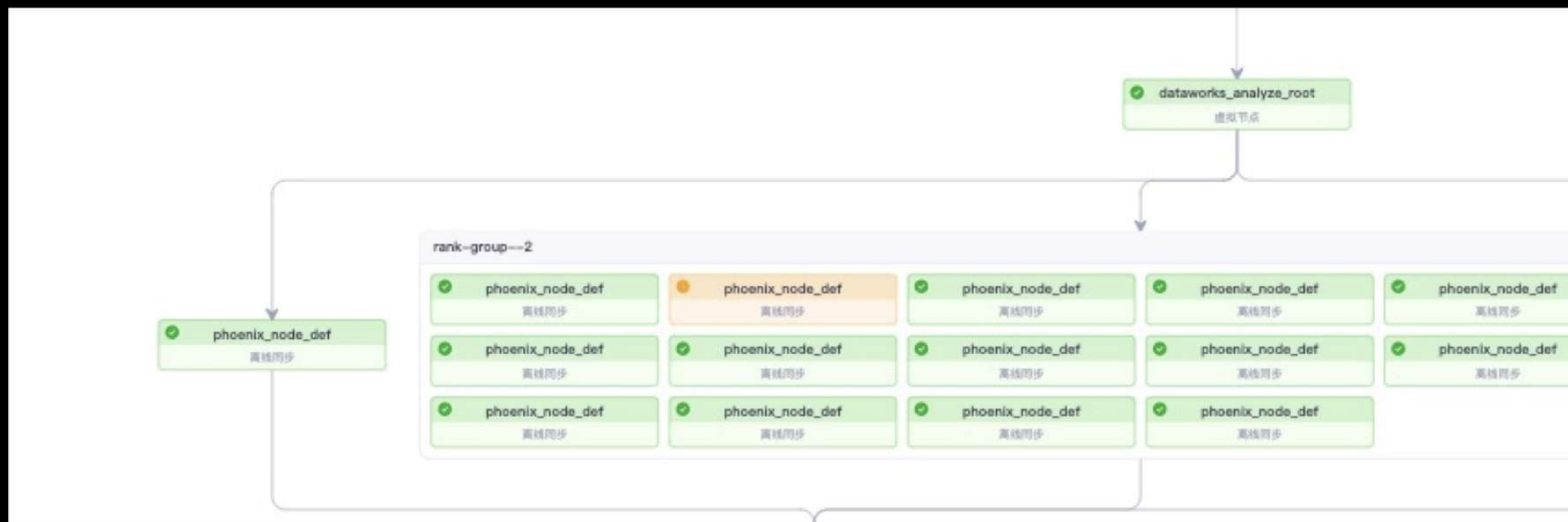
$$NodeX = InnerX + GroupX$$

$$NodeY = InnerY + GroupY$$



布线策略优化

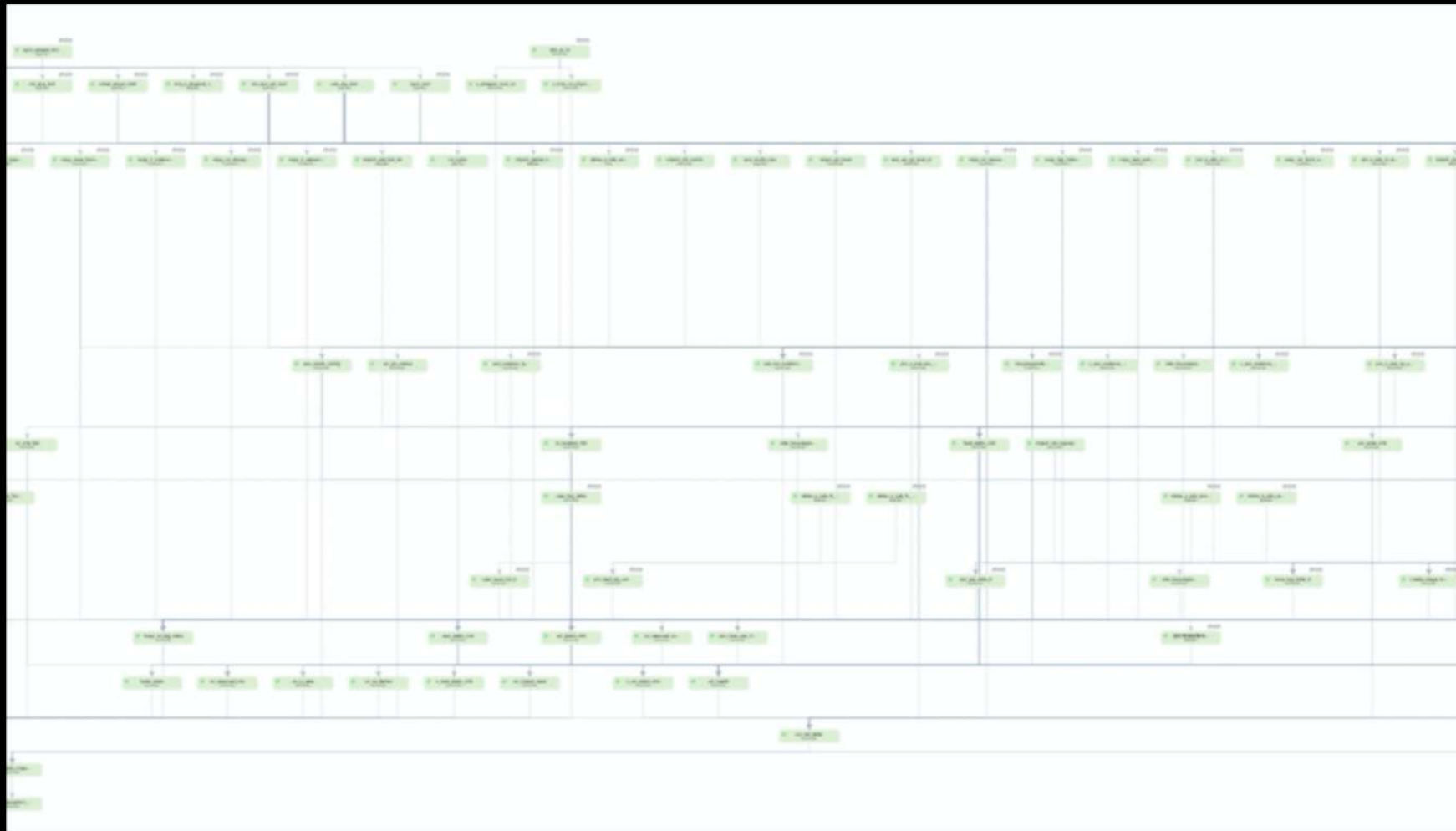
- 替换为Dagre布局，尽量避免边交叉
- 线的路由算法，拐点路由计算
- 直角改为圆角，增加方向性区分



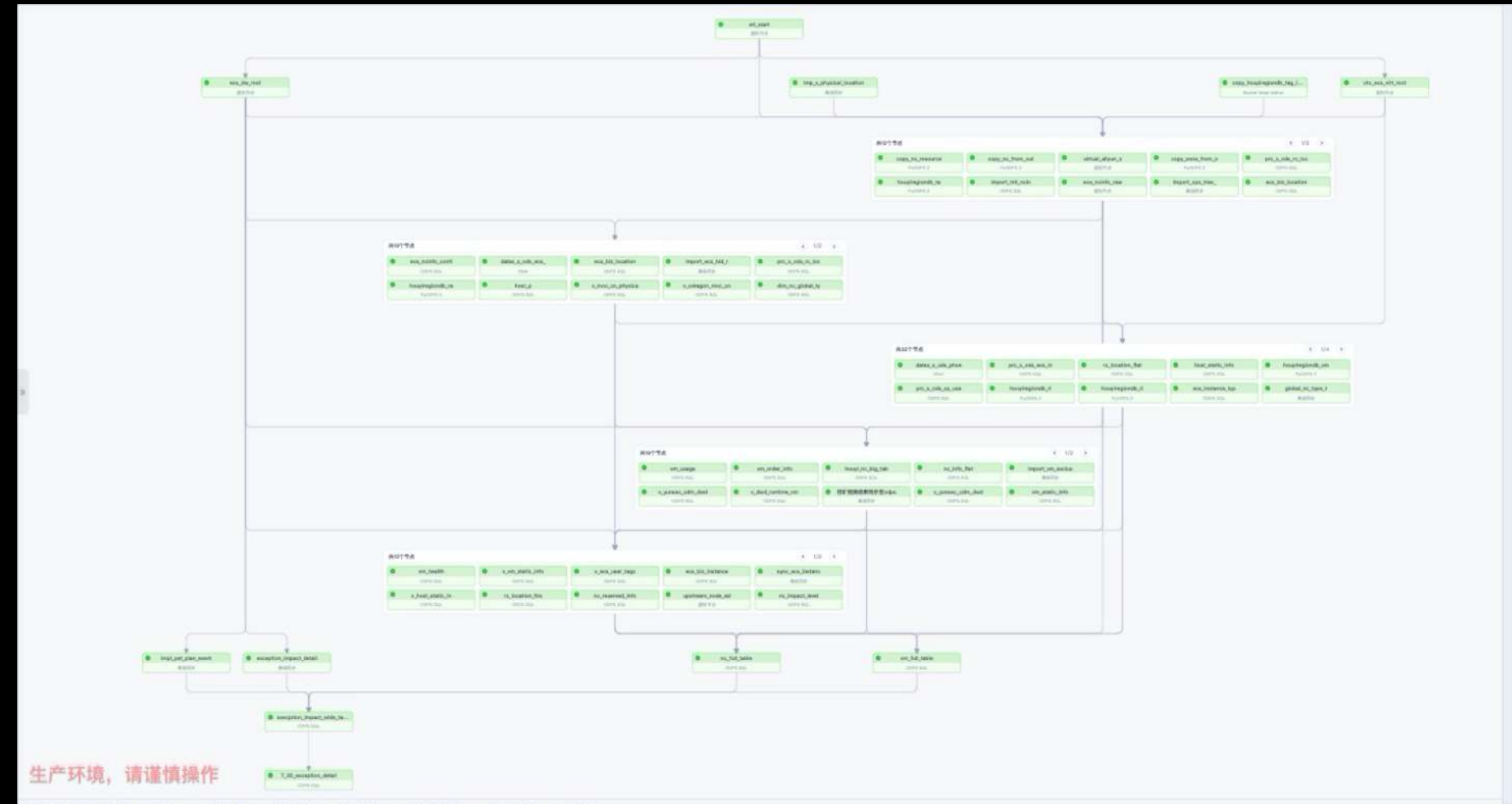
布线策略优化

- 1. 层级分割**：根据所有外部节点（去除内部节点）的y坐标，进行分组，划分层级。
- 2. 边界计算**：根据该层级所有节点的y坐标已经高度，计算出层级区域的界限。
- 3. 拐点确认**：根据层级界限，以及设置的gap，计算出层级之间拐点的Y坐标。
- 4. 绘制曲线**：所有曲线的拐点，按照曲线所处层级的拐点Y坐标绘制。如果是跨多个层级的边，按照最后一个层级的拐点绘制（原因：拐点离终点越近，越容易跟踪线的来源）。

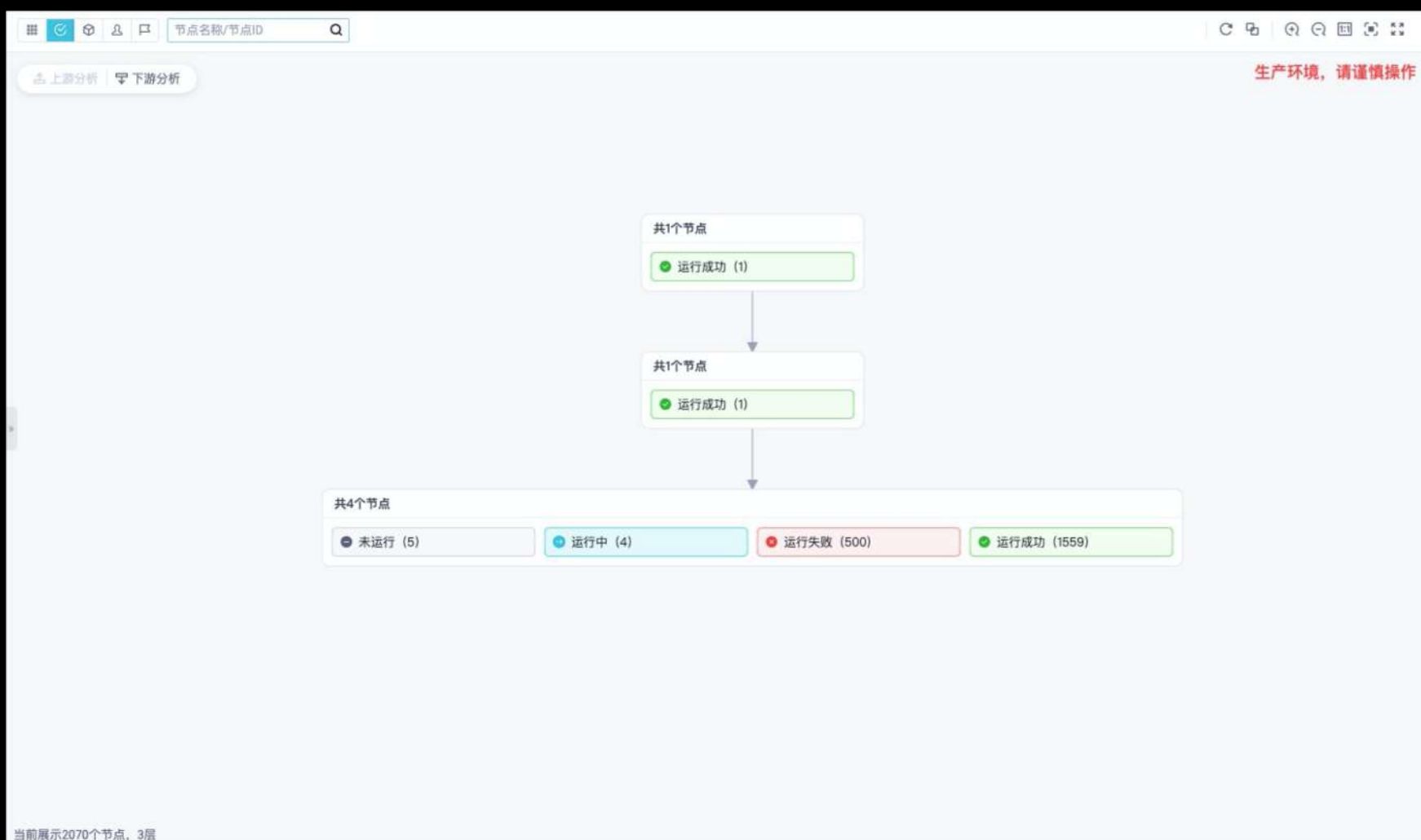




原始布局



Dagre + Grid布局融合

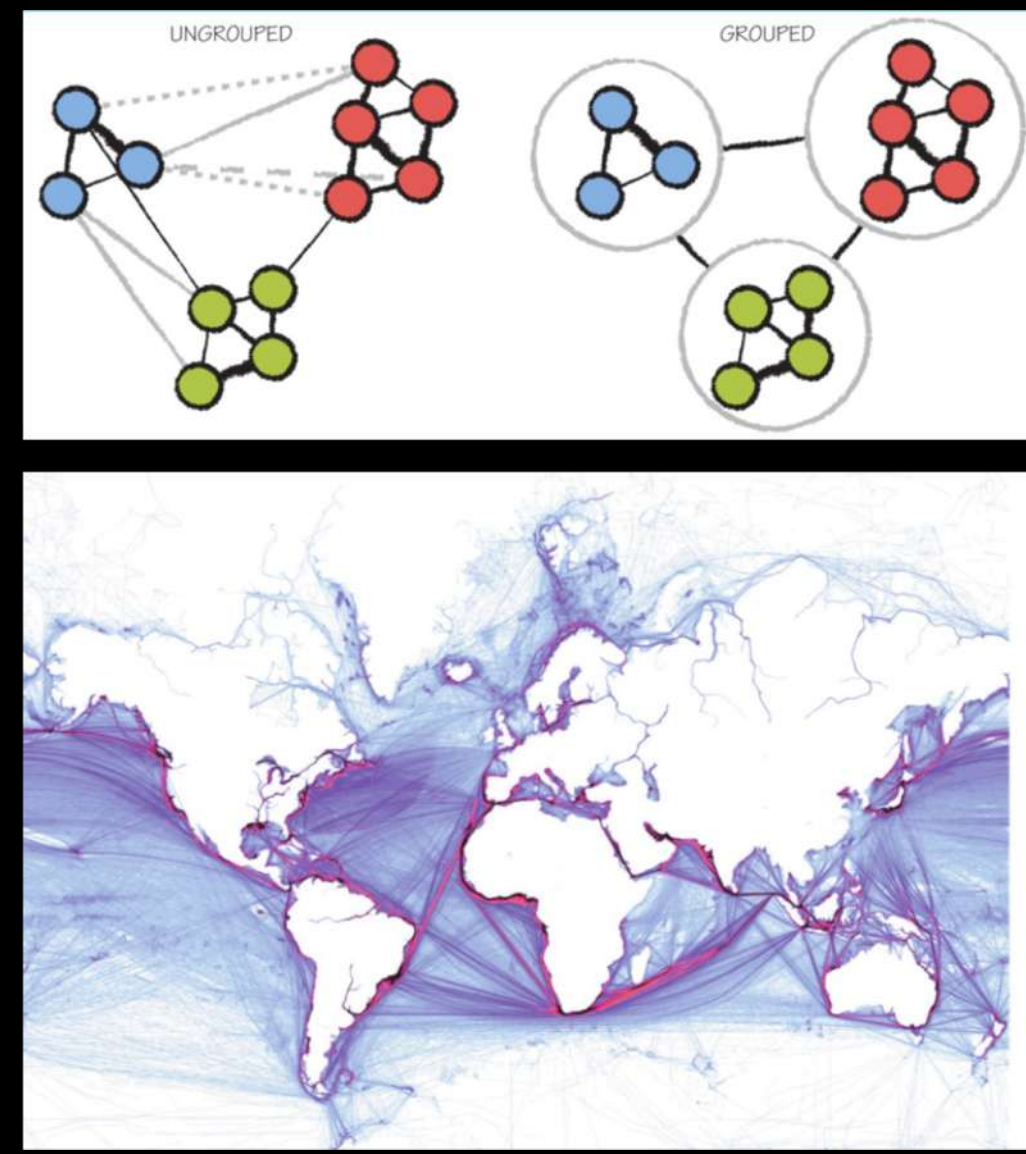


内容聚合

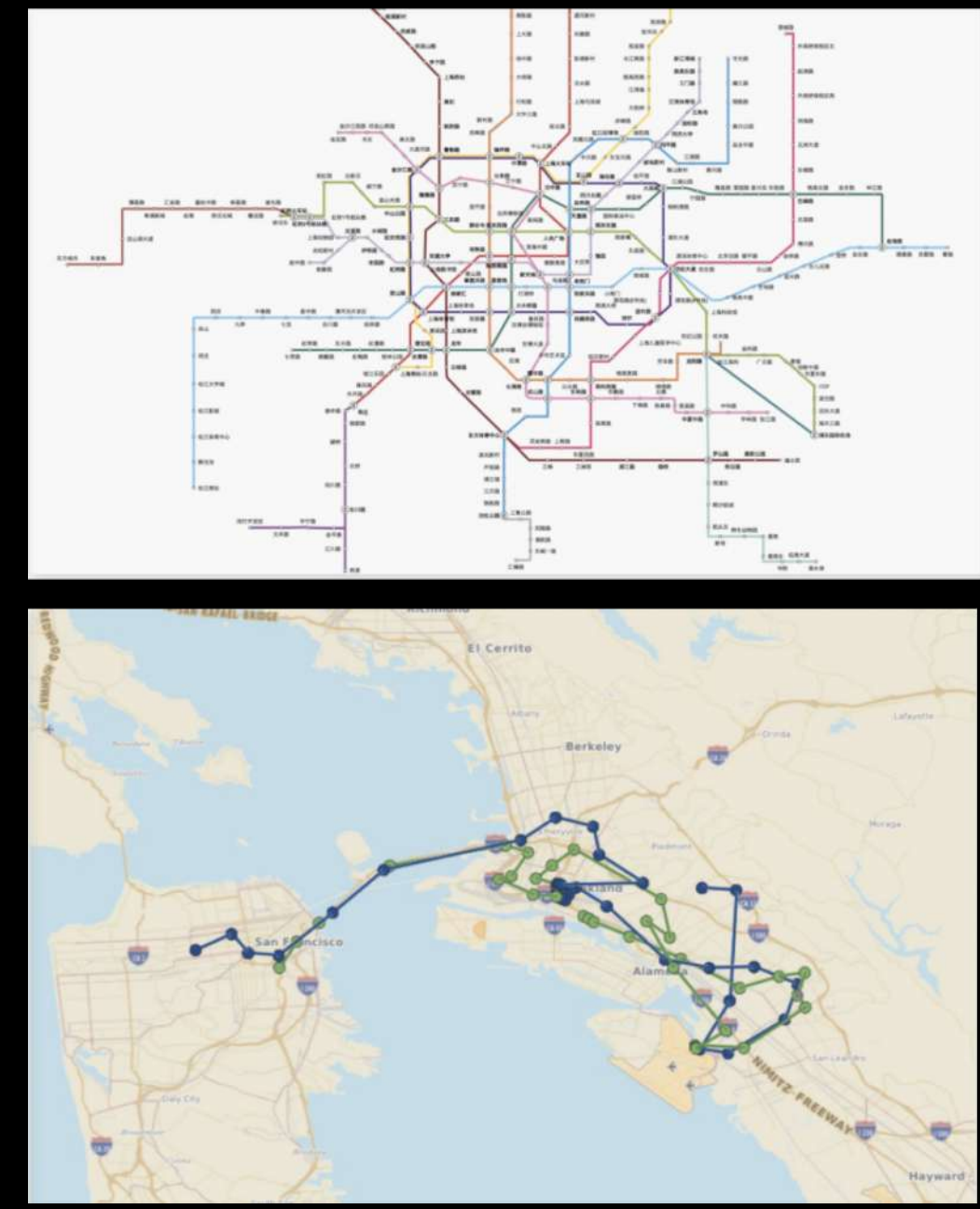


节点翻页

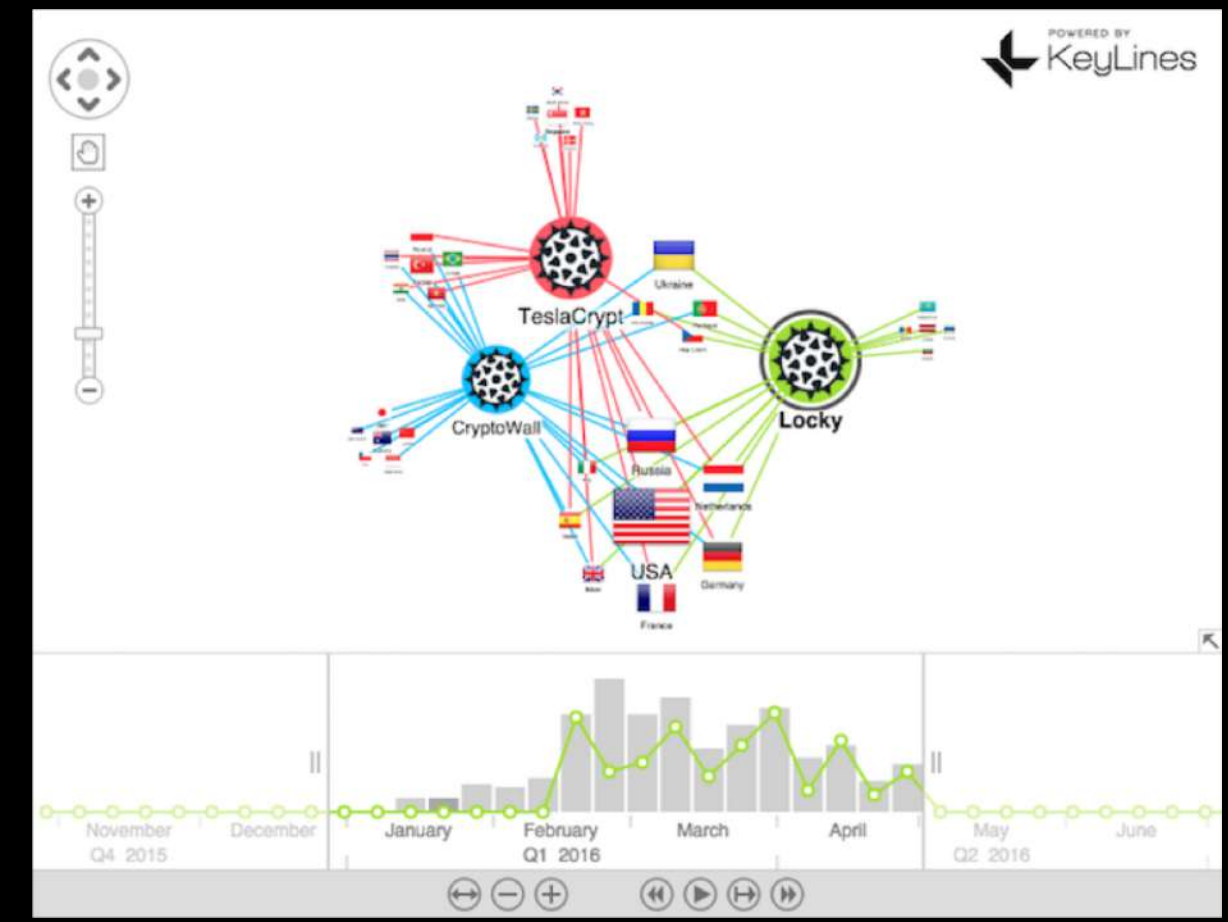
分组分析



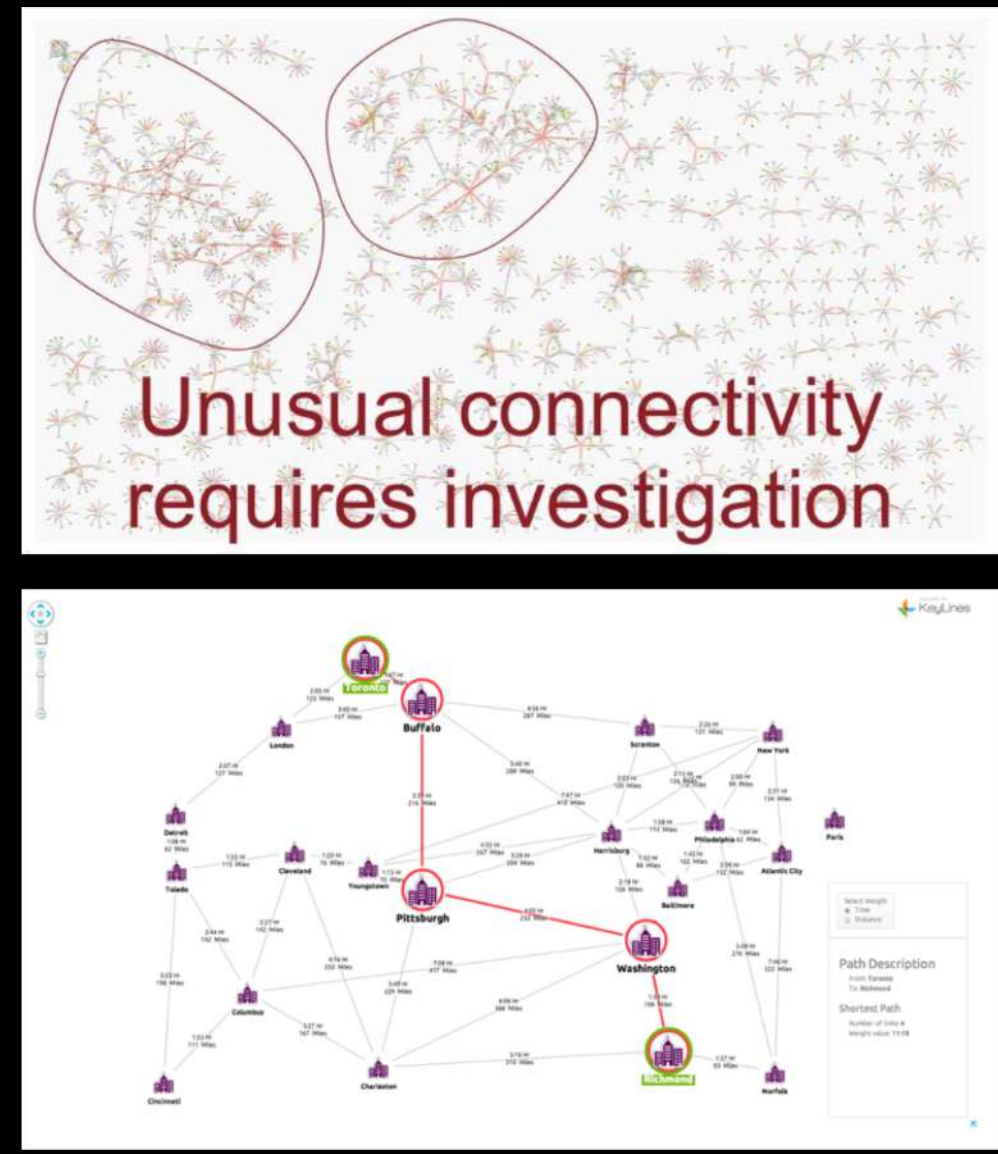
地理空间分析



时序分析



算法分析

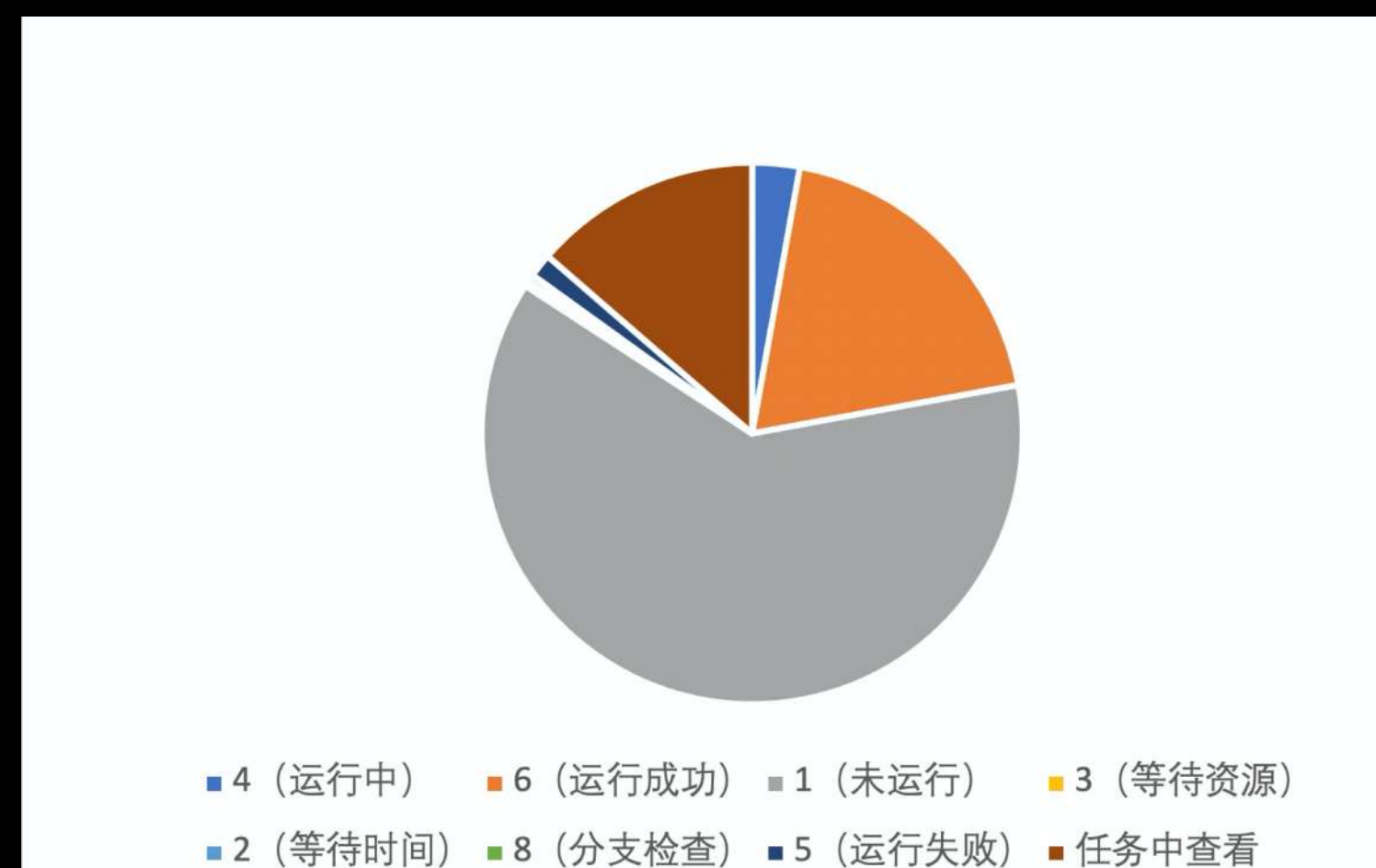


• 背景：

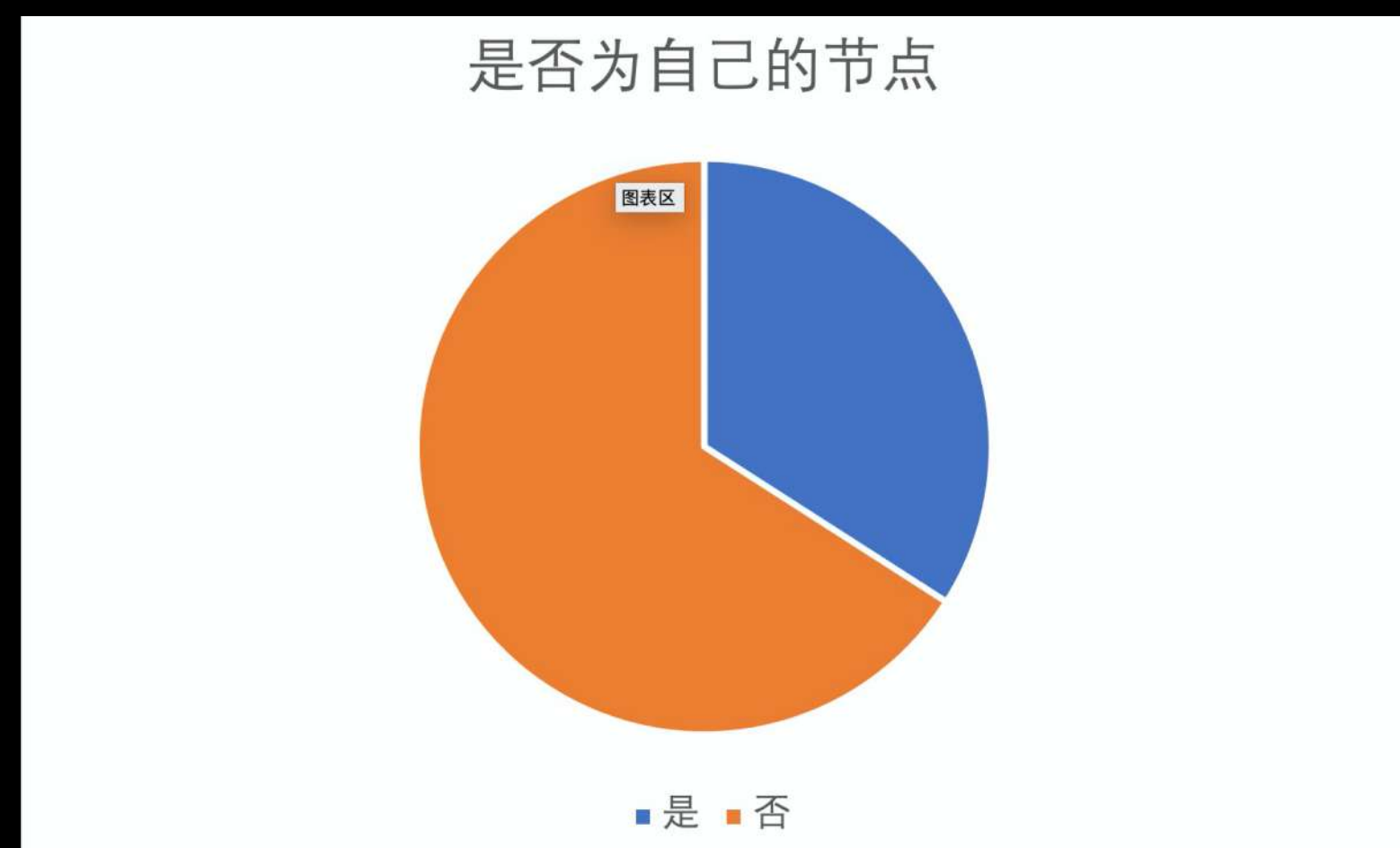
- 用户在什么情况下会点击dag的展开上下游，对应用户想要解决的问题是什么？

• 业务猜想验证：

- 展开下游。用户会针对（自己owner）进行展开下游操作，以判断节点对下游的影响
- 展开上游。用户会针对（错误/未运行）节点进行展开上游操作，去定位上游的阻塞节点

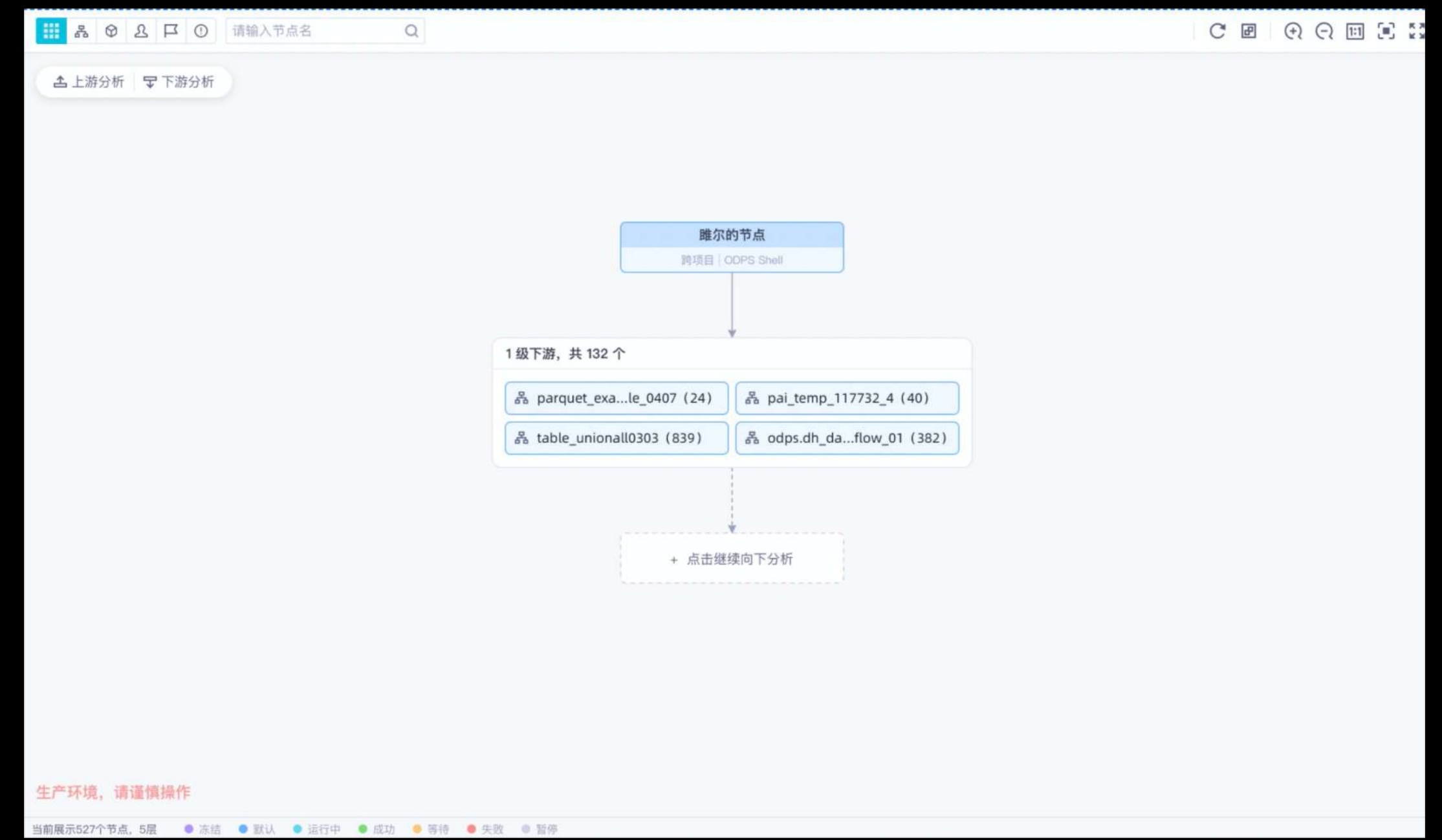
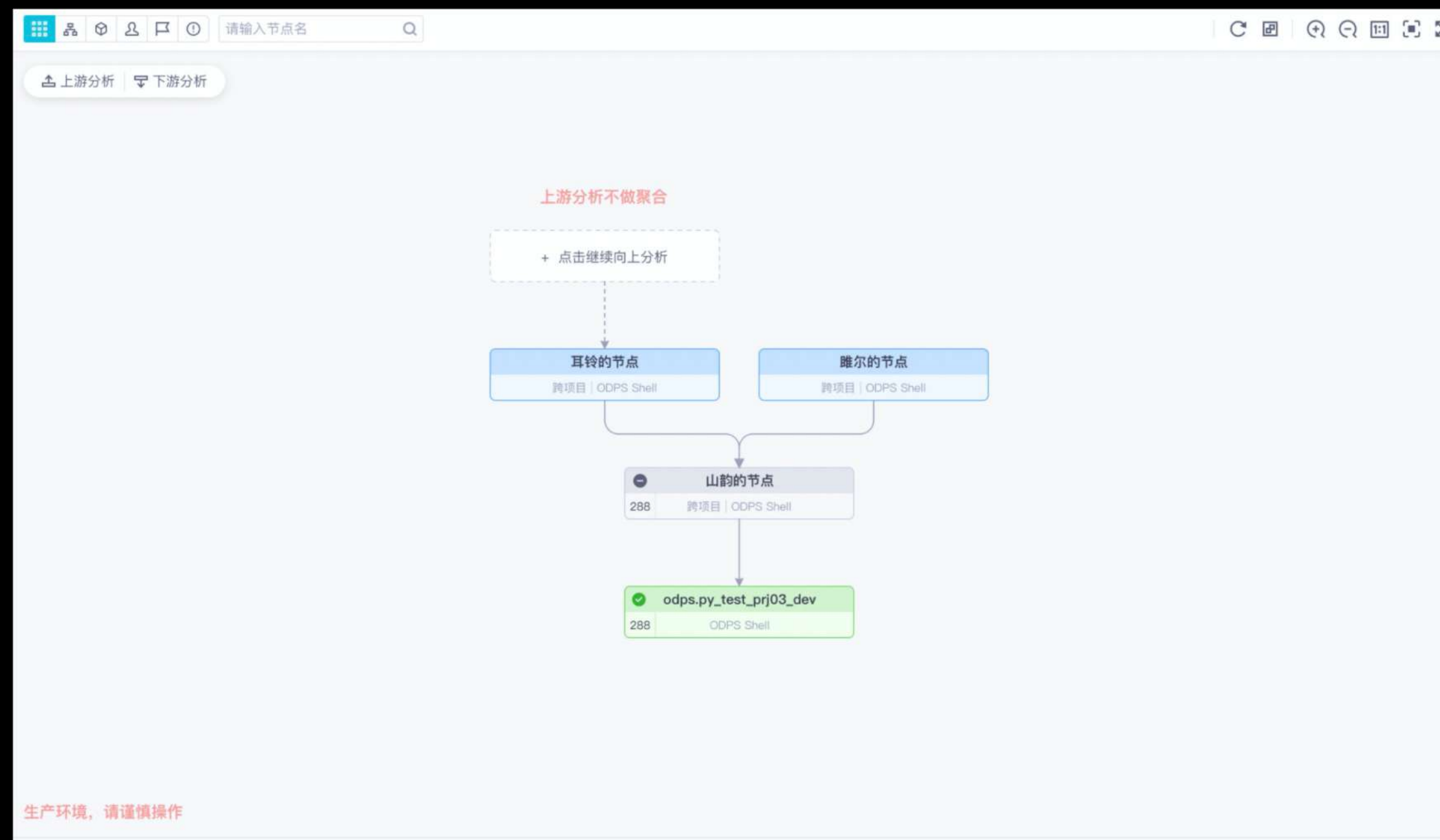


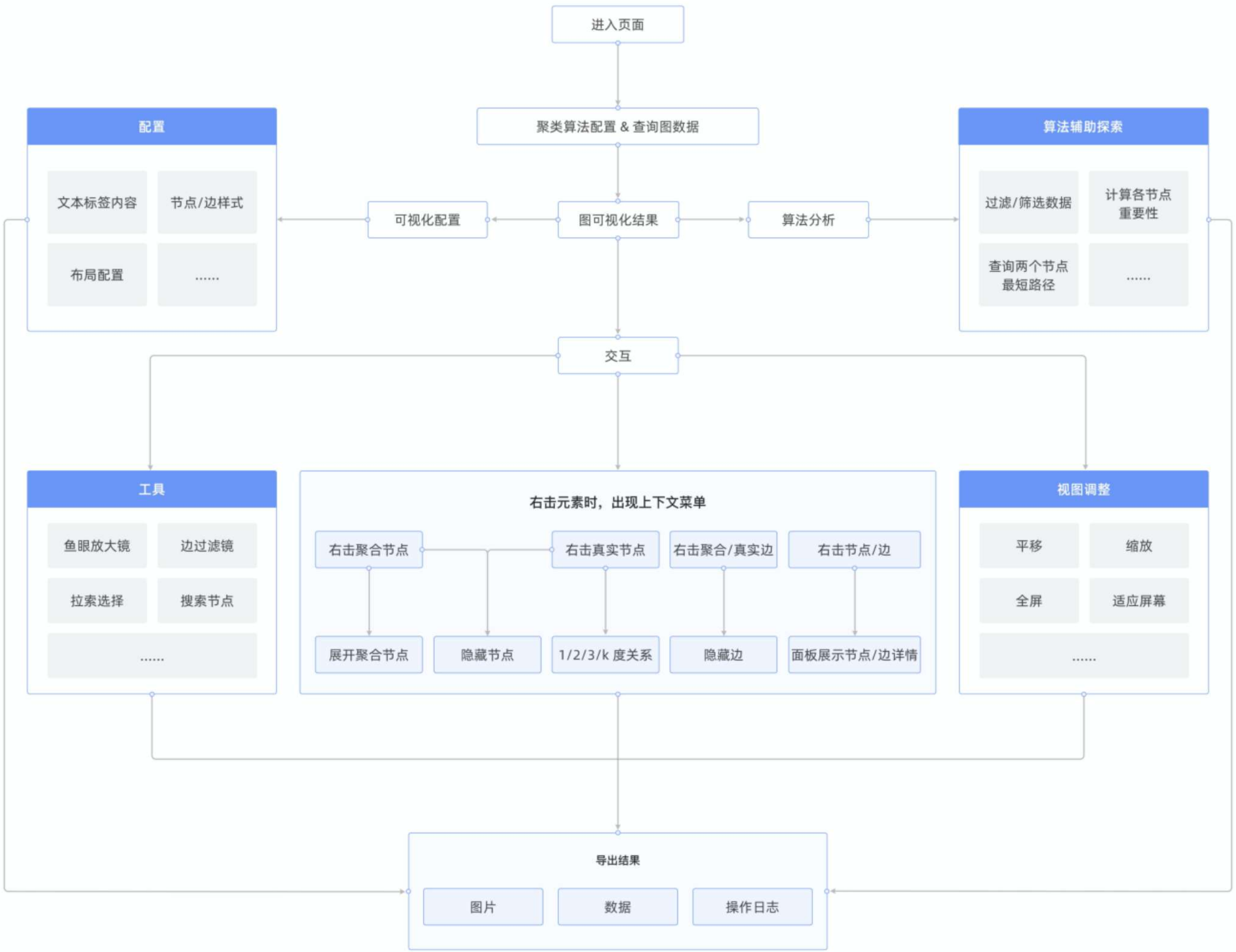
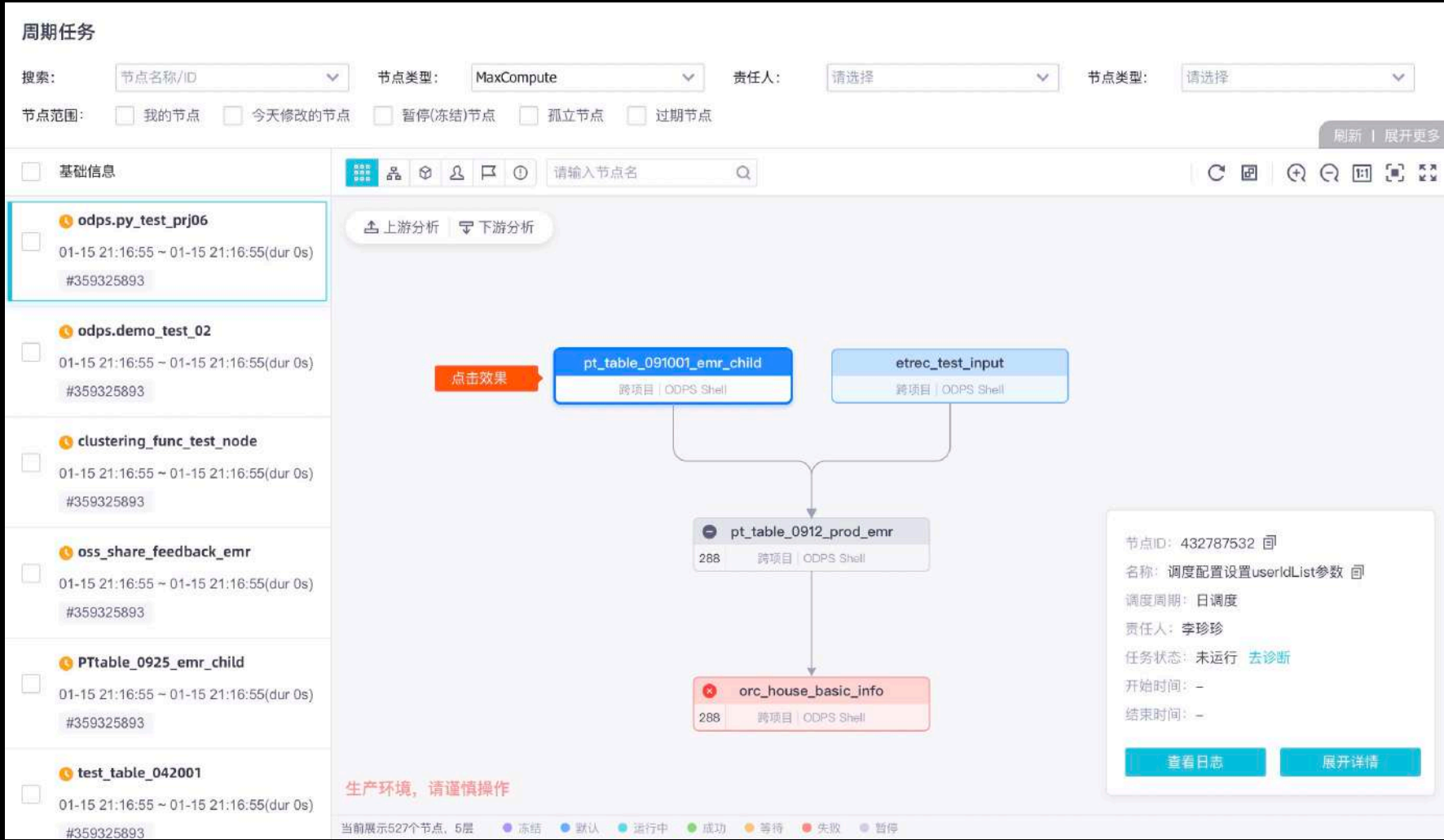
查看上游

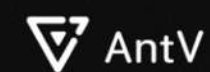


查看下游

- 上游链路分析。针对未运行实例展示阻塞链路
- 下游影响分析。查看节点的下游影响面，提供责任人、项目等多维度视角。







AntV 图可视分析 解决方案

V I S U A L I Z A T I O N

A N A L Y T I C S

S O L U T I O N

O F A N T V

阿里巴巴集团
蚂蚁集团

2020.11.22



图可视化解决方案 云安全

V I S U A L I Z A T I O N

F O R

C L O U D

S E C U R I T Y

阿里巴巴集团
蚂蚁集团

2020.11.22



图可视化解决方案 图数据库

V I S U A L I Z A T I O N

F O R

G R A P H

D A T A B A S E

阿里巴巴集团
蚂蚁集团

2020.11.22



AntV 图可视分析 设计指引

D E S I G N

G U I D E L I N E

F O R

G R A P H

V I S U A L I Z A T I O N

阿里巴巴集团
蚂蚁集团

2020.11.22



图可视化解决方案 知识图谱

V I S U A L I Z A T I O N

F O R

K N O W L E D G E

G R A P H

阿里巴巴集团
蚂蚁集团

2020.11.22



图可视化解决方案 企业风控

V I S U A L I Z A T I O N

F O R

E N T E R P R I S E

阿里巴巴集团
蚂蚁集团

2020.11.22



图可视化解决方案 性能优化

P E R F O R M A N C E

O P T I M I Z A T I O N

V I S U A L I Z A T I O N

阿里巴巴集团
蚂蚁集团

2020.11.22



Thanks