

NebulaGraph + LLM 处理风控知识图谱的 探索和实践

余毅航 (盐粒/Yanli)

VESoft 工程师

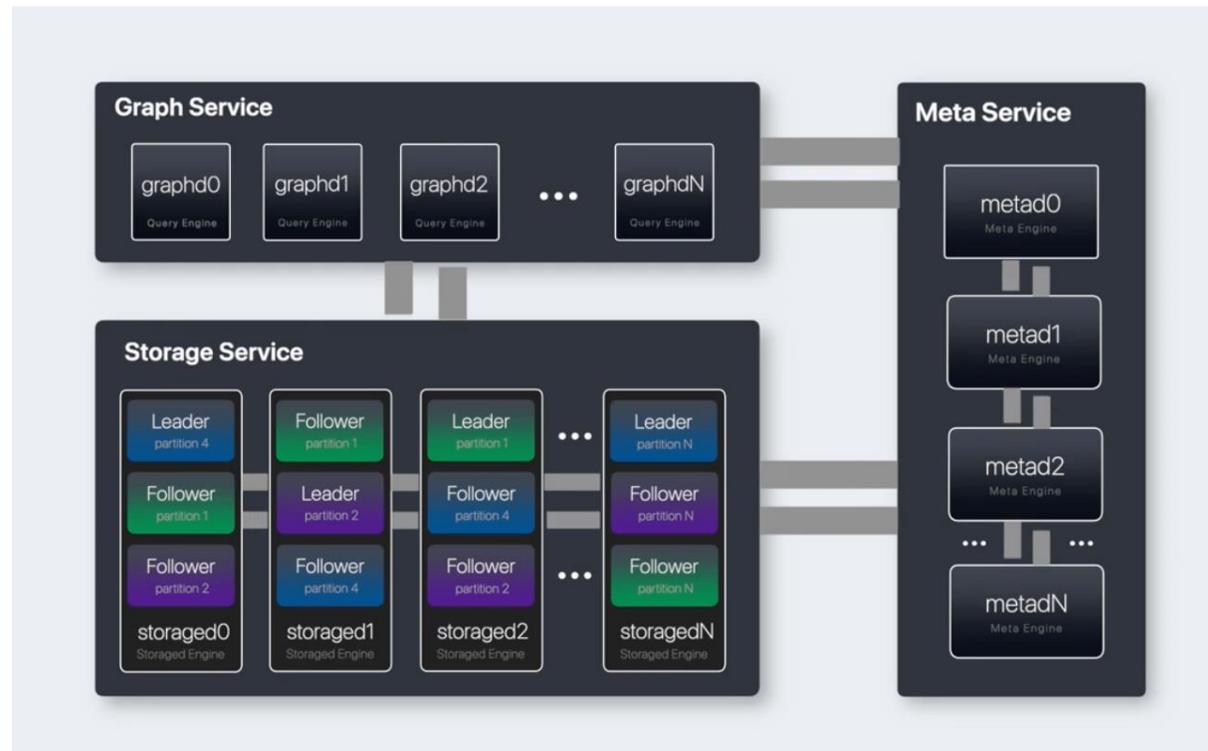
DataFunSummit # 2024





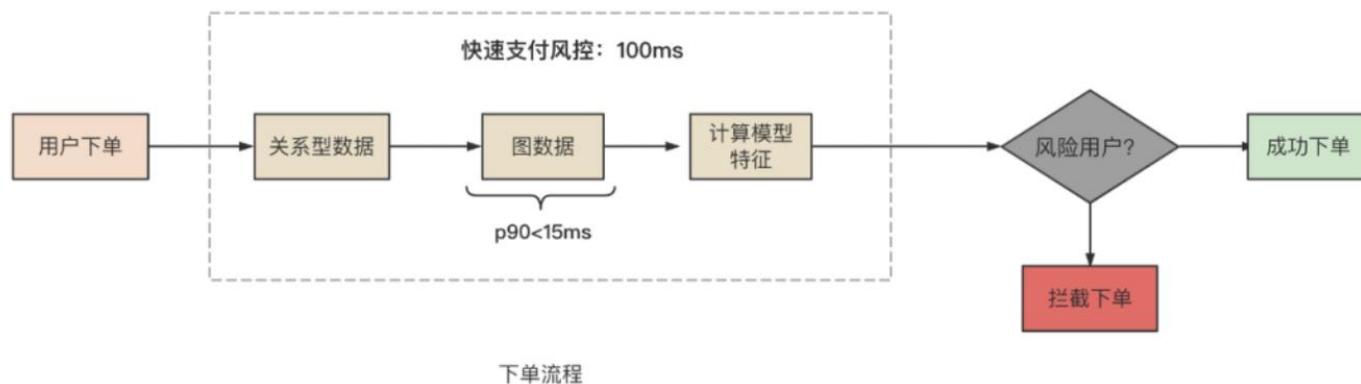
超大规模图数据库，毫秒级延迟。

开源、分布式、可扩展、极速

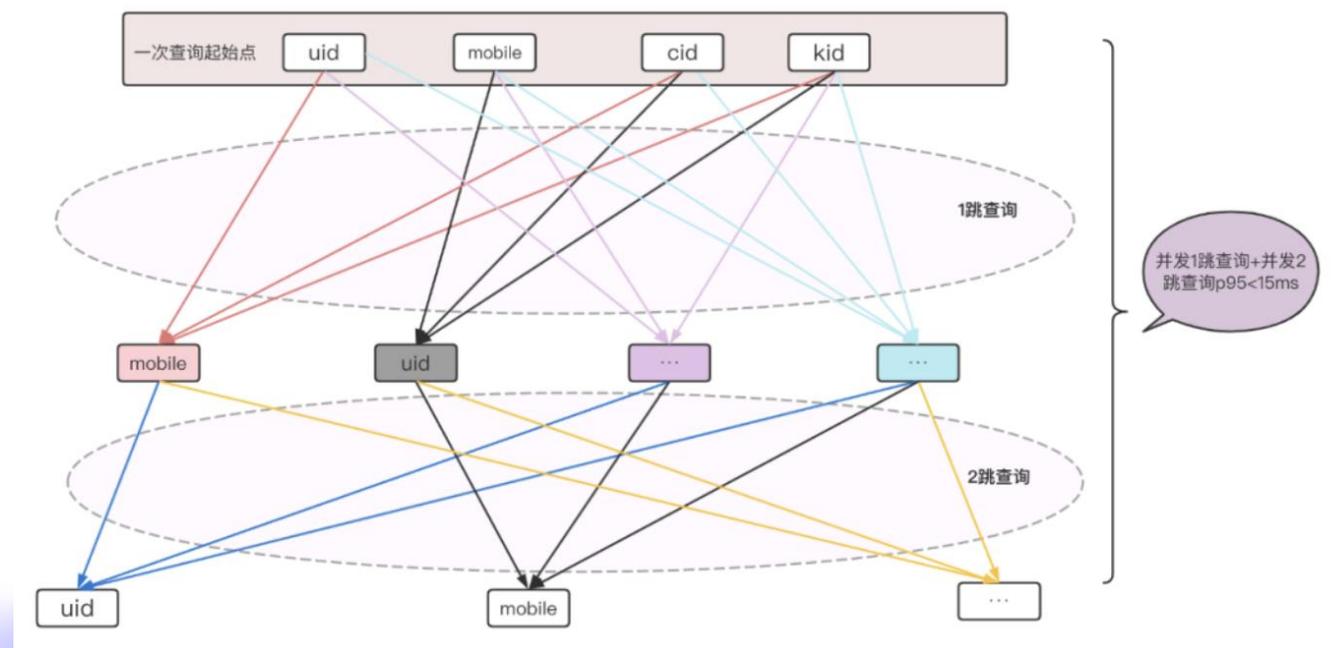


在 NebulaGraph 上构建 应用风控图谱的成熟实践

Best Practice 【金融实时风控场景】



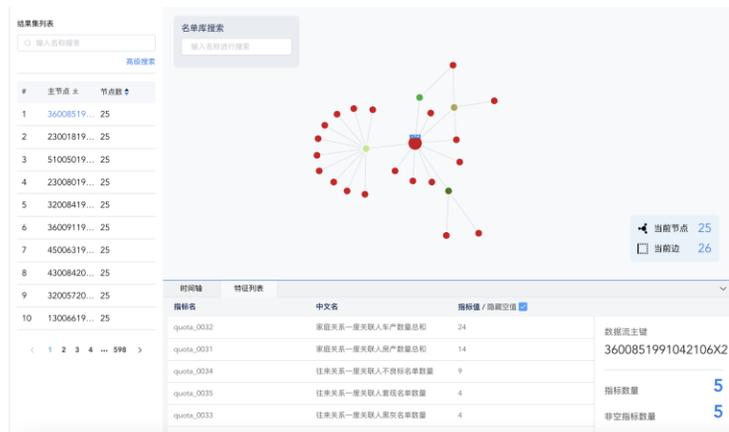
用户下单时，会进入一个快速风控的阶段：通过基于关系型数据库和图数据库的规则进行模型特征计算，来判断这个用户是不是风险用户，要不要对该用户进行下单拦截（实时反欺诈）。



已知某个 uid 是犯欺团伙的一员，根据图关联来判断跟他关系紧密的用户是不是存在欺诈行为。

为了避免影响正常用户的下单流程，风控阶段需要ms级响应，因此对图查询的性能要求非常高 (P95 < 15 ms)。基于 [NebulaGraph](#) 构建了百亿级的反欺诈图；

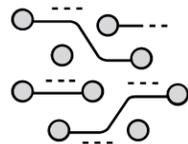
实时查询与离线计算的融合架构



命中列表线索探查



业务系统



Kafka



其他风险客户
以及关联线索



实时构图流程
实时指标计算

离线数据处理

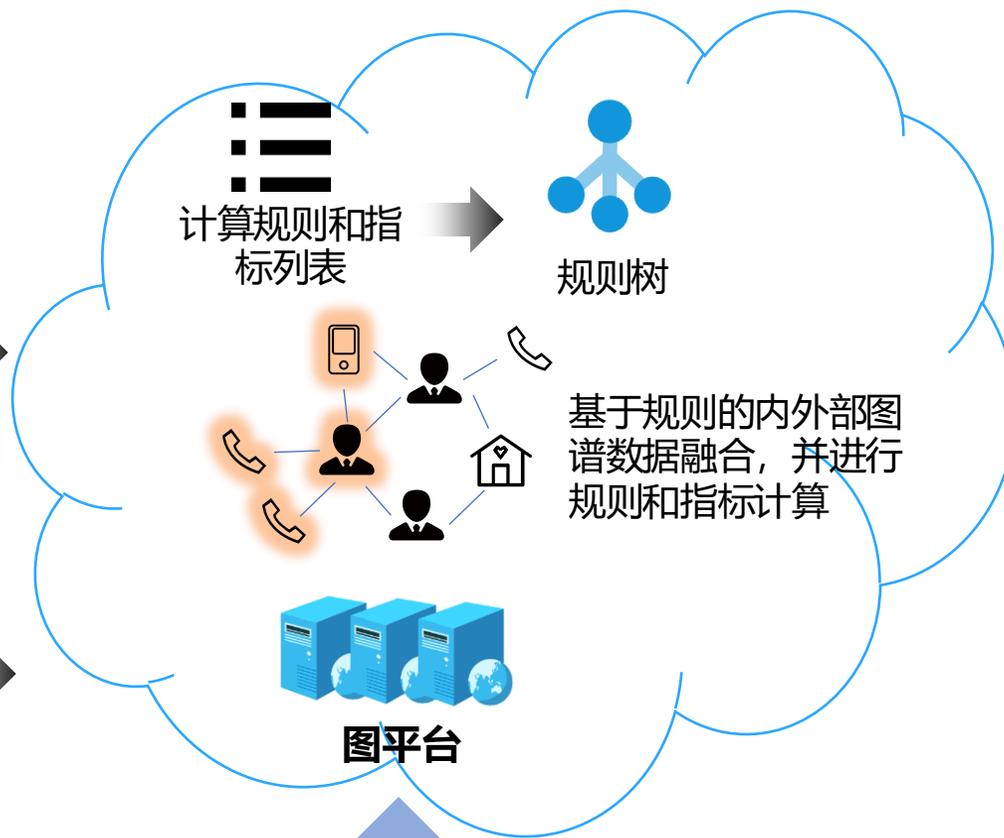
采集落入Hive表

清洗标准化

数据预处理

获取潜在客户号
进行实体对齐

T+1离线构图



常见的实时欺诈分析方法--图指标

分析申请人的单位、家庭、设备、LBS等各维度关联人子图的质量水平；2、通过模式识别特定的网络结构，从而识别资料异常的申请人，或申请人关联的特定客群。

维度

单位类	单位名称、单位地址、单位电话
家庭类	住宅地址、住宅电话、车牌号
设备类	Ip地址、mac地址、IMEI串号、设备ID
LBS类	身份证影像坐标、亲访照影像坐标
联系方式	手机、电子邮箱、紧急联系电话、卡邮寄地址
推荐方式	推荐人、营销员、代理人
亲属	父母、子女、配偶、岳父母、兄弟姐妹
其他往来关系	担保人、朋友、资金往来、同还款人...

资料异常

本人单位类信息为他人家庭类信息
本人家庭类信息为他人单位类信息
本人电话同时为他人电话

特定客群

同一小时内进件身份影像坐标相同
同一小时内进件ip地址相同
同一推荐方式且相同工作单位
同一推荐方式且有家庭关系

客户标签

黑名单	逾期
申请拒绝	止付
欺诈拒绝	申请通过

还款能力

年收入
负债
房产数量
车产数量

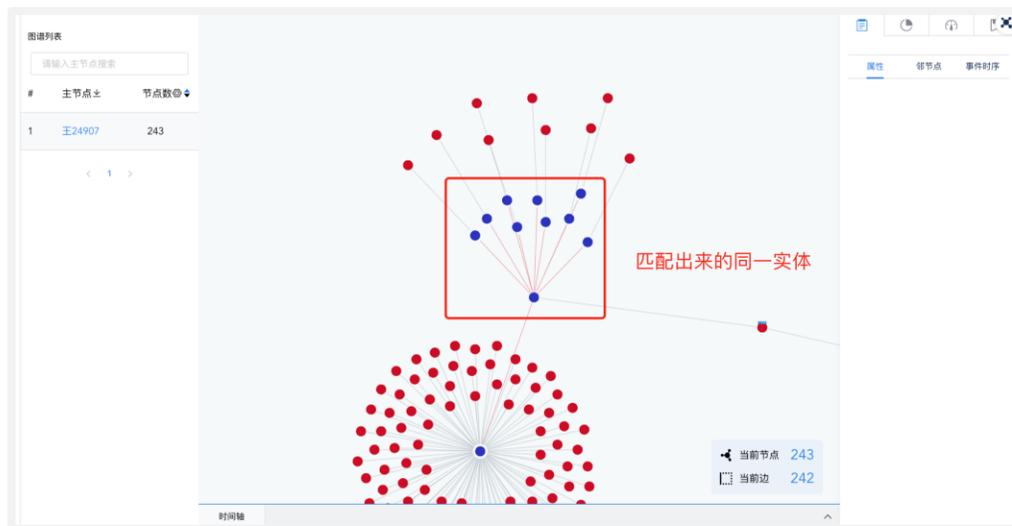
还款风险

人行查询次数
银行多头申请笔数
延滞月份
拦截交易笔数

快速构建各类业务规则和指标---规则

平台的分析模块中，包含了快速图谱规则定义的能力，可以快速将各类的业务逻辑进行转换

1. 知识定义中切换的规则模块
2. 新建规则，在交互页面中进行规则的构建
3. 在指标或作业流程中运行1或多个规则，如命中规则则可以在结果列表中进行探查



名称	类型	对应实体类型	对应实体属性	相似度
模糊	地址	地址	地址	大于0.7
模糊匹配指标0901	地址	地址	地址	等于0.0
模糊匹配模型版本...	地址	地址	地址	小于0.1
模糊匹配指标	地址	地址	地址	等于0.7
地址介于	地址	地址	地址	小于等于
公司	公司名称	公司	公司	介于(0.5
地址2	地址	地址	地址	大于等于
地址	地址	地址	地址	大于等于

地址 V3

数据权限: 公开可编辑
类型: 地址
对应实体类型: 地址
对应实体属性: 地址
相似度分数: 大于等于0.5
返回分数TOP N结果: 100

模糊匹配条件:
模糊匹配模型: 模糊匹配指标0901
选择版本: V9

快速构建各类业务规则和指标---指标

平台的分析模块中，包含了快速指标定义的能力，可以基于规则匹配的结果或者社区切分的结果进行计算分析

1. 知识定义中切换指标模块
2. 新建指标，选择基于规则的实体指标或基于社区切分结果的子图指标
3. 在作业流程中运行N个指标计算，可以选择实时计算或者离线计算，会将计算结果以列表形式进行展现

类型	使用场景
实体指标	利用行业专家的强规则对规则子图进行指标计算，可应用于信贷业务审核、骗保风险预测等场景
子图指标	面向算法子图，针对社区进行的指标计算，展示社区之间的区别和特点

新建实体指标

类型: 实体指标

规则: 人-公司-人(模糊匹配模型)

支持实时: 是 否

权限: 公开

指标列表: 共1个

- 名称: 请输入指标名称
- 中文名: 请输入指标中文名
- 算子: 总数
- 计算主体: 个人2 + 筛选条件
 - 属性筛选: 欺诈拒绝代码 等于 1

取消 保存

新建子图指标

类型: 子图指标

权限: 公开

指标列表: 共1个

- 名称: 请输入指标名称 (指标名不能为空)
- 中文名: 请输入指标中文名 (请输入中文名)
- 算子: 总数
- 计算主体: 个人 + 筛选条件

取消 保存

快速构建各类业务规则和指标---指标

实体指标计算、
子图指标计算、
实时指标计算

指标计算结果展示，包含计算数值以及计算口径

导图算子

- 离线导图
- 实时导图

分析算子

- 图规则匹配
- Louvain
- LPA
- 连通图计算
- 实体指标**
- 子图指标**
- 子图实体指标
- 模糊匹配分团

实时图分析算子

- 实体指标**

数据处理算子

- SHELL
- 数据同步

实体指标-计算非模糊匹配
本体: 团伙反欺诈

输入名称搜索

下载子图 保存子图 添加节点 添加连线 力导向布局 树型布局 网格布局 布局参数 撤销 重做 删除 清空画布 刷新 社交领袖 社区发现 清除算法

结果集列表

#	主节点	节点数
1	15004819...	14
2	37001219...	14
3	13001019...	14
4	15002820...	12
5	45008019...	12
6	51001220...	12
7	23008019...	12
8	22004519...	12
9	15001719...	12
10	51005019...	12

名单库搜索

输入名称进行搜索

当前节点 14
当前边 13

指标名	中文名	指标值 / 隐藏空值	数据流主键
quota_salary	计算同一公司的平均年薪	3424232	150048199604030828
quota_black	计算套现名单	4	

时间轴 特征列表

属性 命中名单 邻节点 关联时序

唯一标识: 150048199604030828

描述字段: 150048199604030828

类型: 个人

进件渠道: 无数据

欺诈拒绝代码: 无数据

客群: 无数据

信用风险等级: 无数据

审批结果: 无数据

公积金社保情况: 无数据

核卡日期: 无数据

申卡时间: 无数据

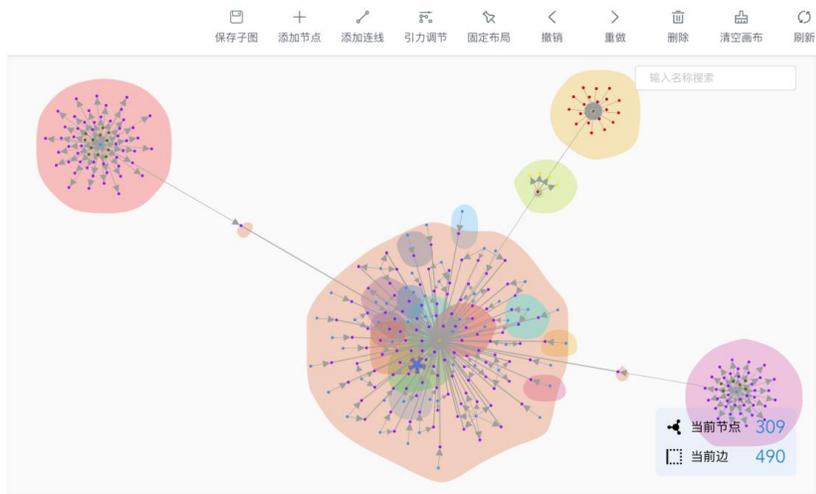
申卡条形码: 无数据

指标数量 2
非空指标数量 2

规则指标+图算法拓展

使用社区切分算法，可以基于客户的关系（如转账，同一IP，设备号，联系电话，相似地址等），将客户切分为紧密社群，紧密社群中的每一个客户彼此之间都有一定的关联强度。通过计算不同维度下的客群指标，来挖掘可疑的个体

例如：客群关联人黑灰名单数量，不良标名单数量，命中特定可疑标签的比例等等，实现对规则**准确率足够但覆盖率不足**的问题



社区切分技术

+

* 名称

* 中文名

* 算子

* 计算主体 ⊕ 添加

+ 筛选条件

属性筛选

近6期账单是否出现过延滞 等于 1 X

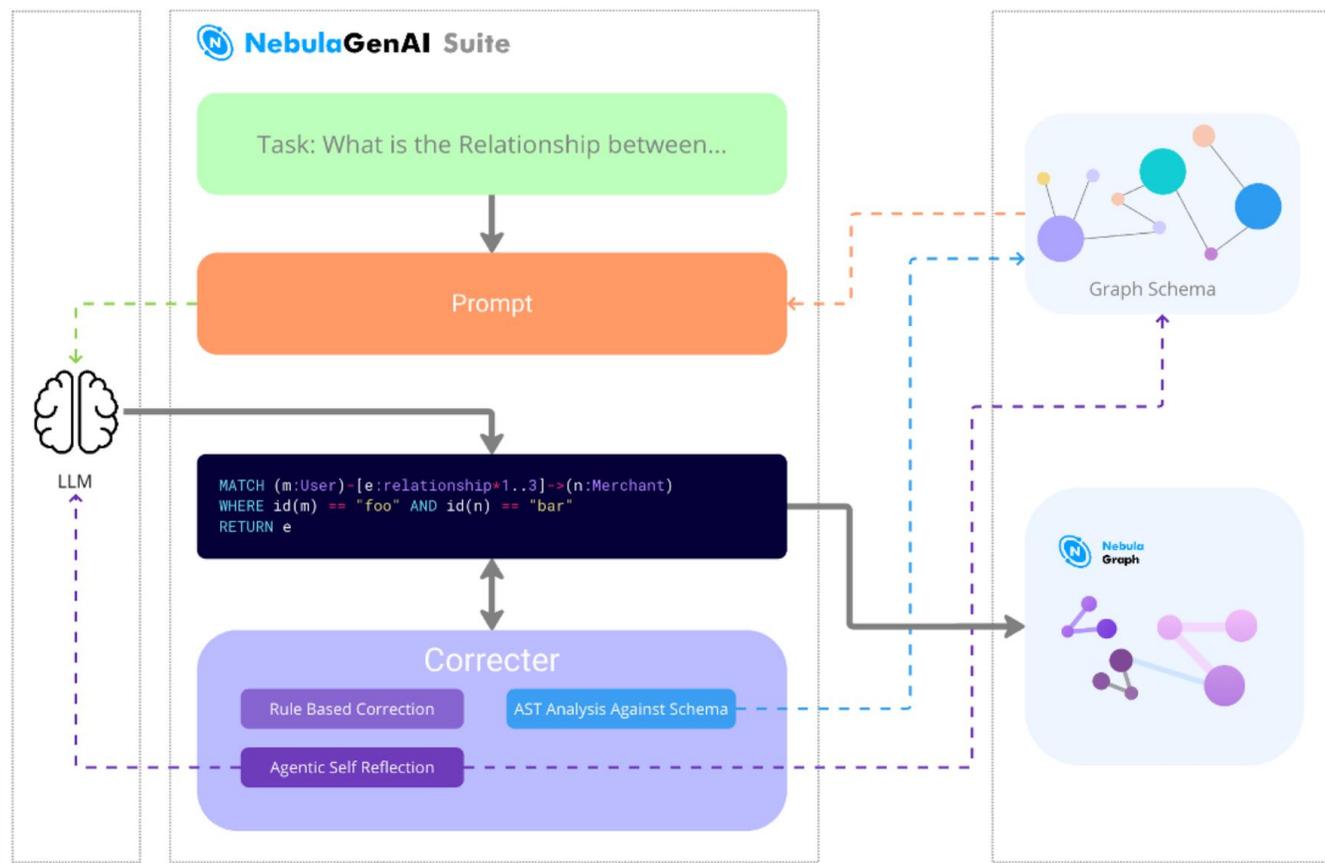
图指标计算技术

LLM 基于图可以做什么？

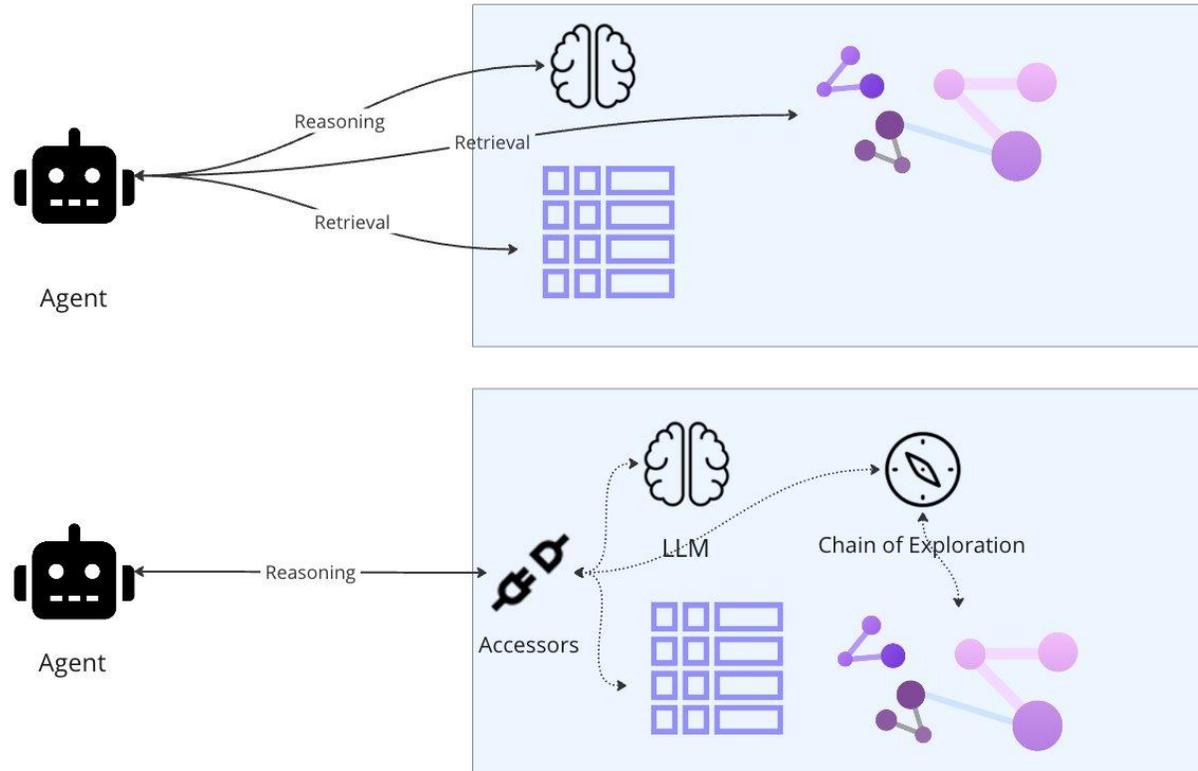
在知识图谱上引入 LLM Agent 的思路

Text to GQL

领域专家可以不需要使用图查询语言
而是直接使用自然语言查询



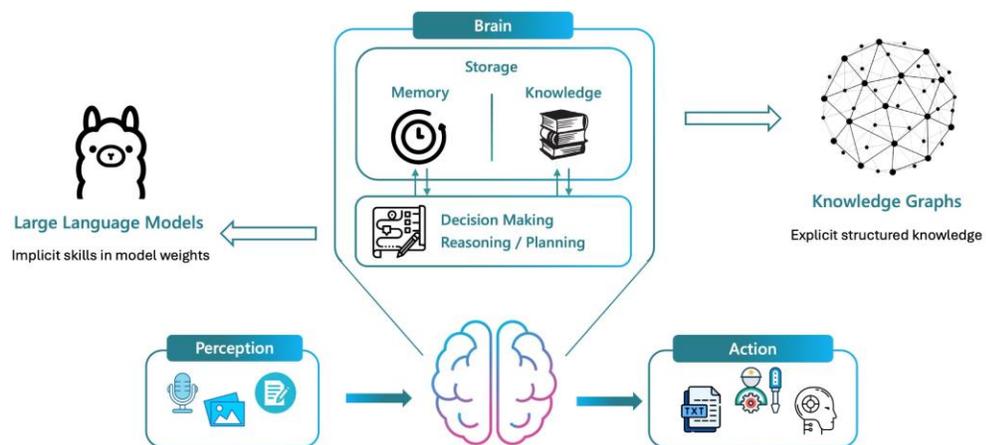
让 Agent 可以理解图谱



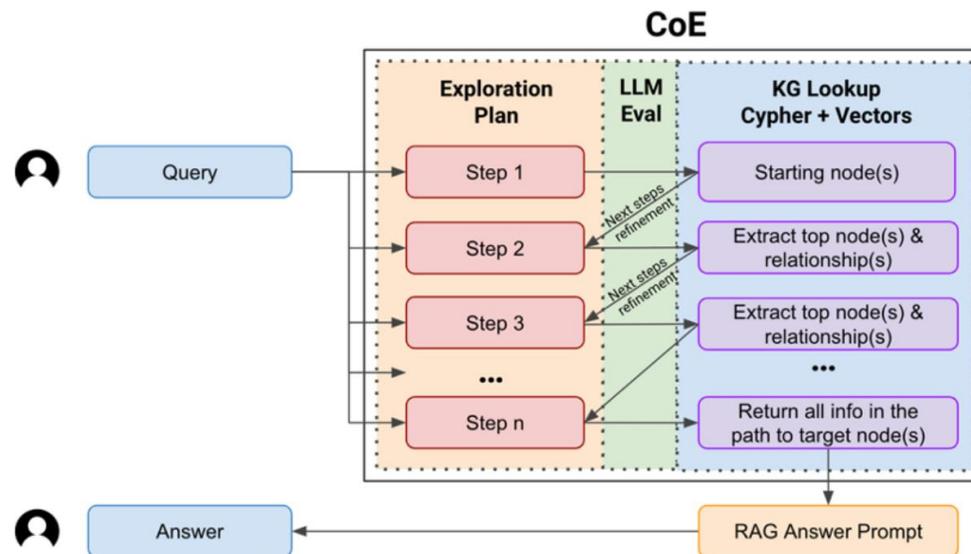
Agentic Reasoning Layer with Hyper-Scale Domain Knowledge and Online Data

Chain of Exploration : 探索链

结合查询和理解的逐步 Agent



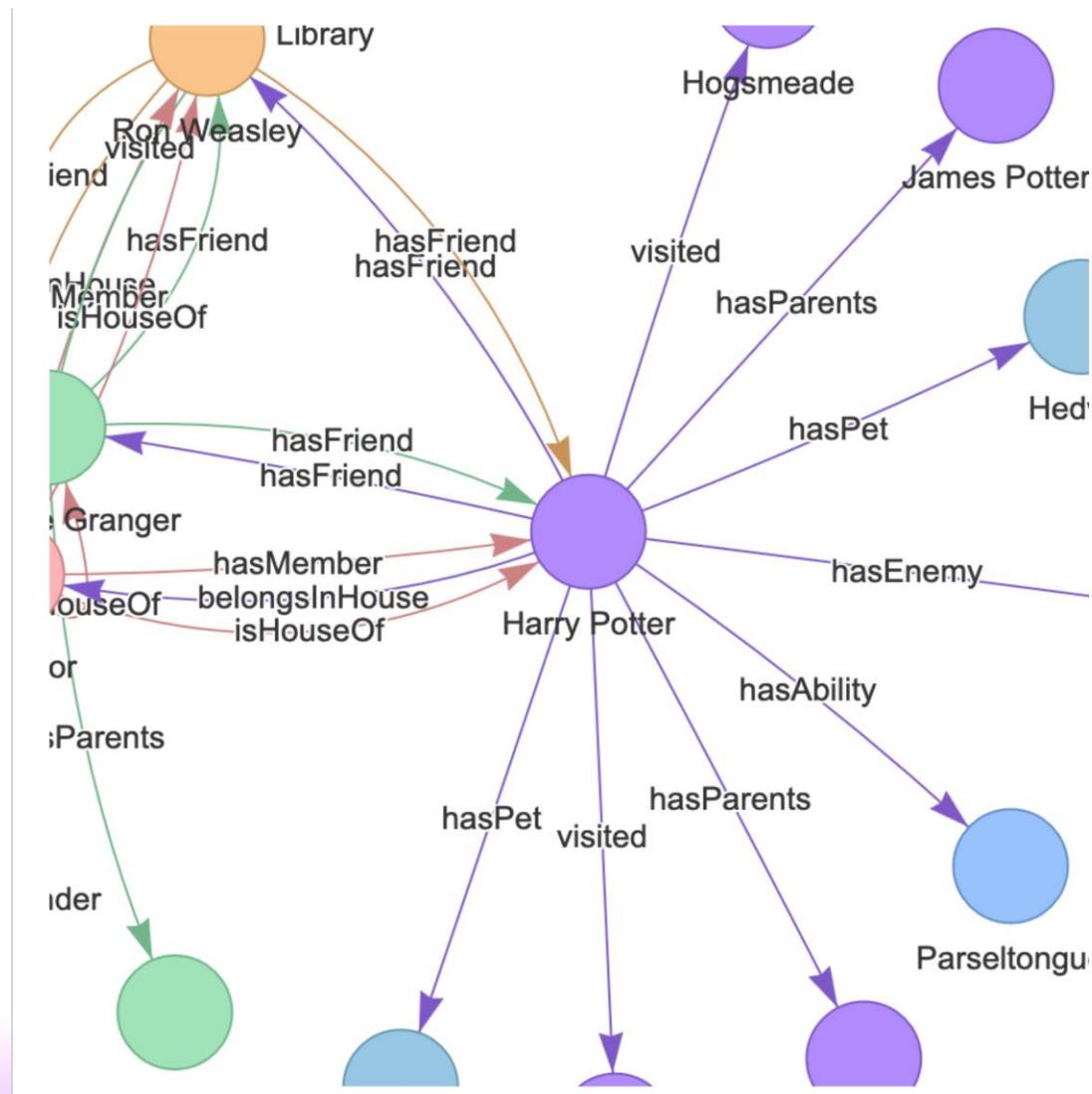
Large Language Models
Implicit skills in model weights



除了结构化数据呢？

从非结构化数据到半结构化知识图谱

从非结构化数据中 构建知识图谱



构建图谱的象限划分

领域特定型：

有明确的点边类型，例如交易流水、人物关系

通用型：

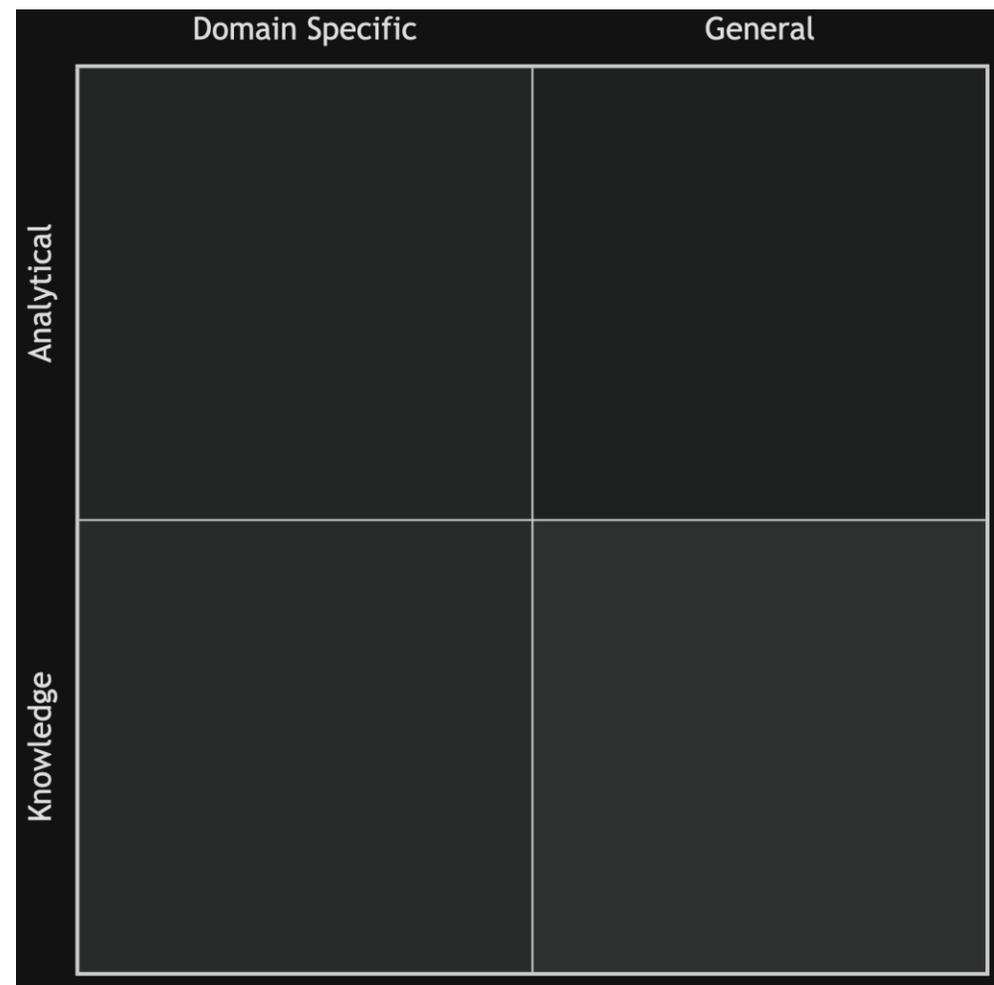
无明确类型，例如主谓宾三元组

分析型：

注重结构化查询的能力，使用 GQL 查询语言

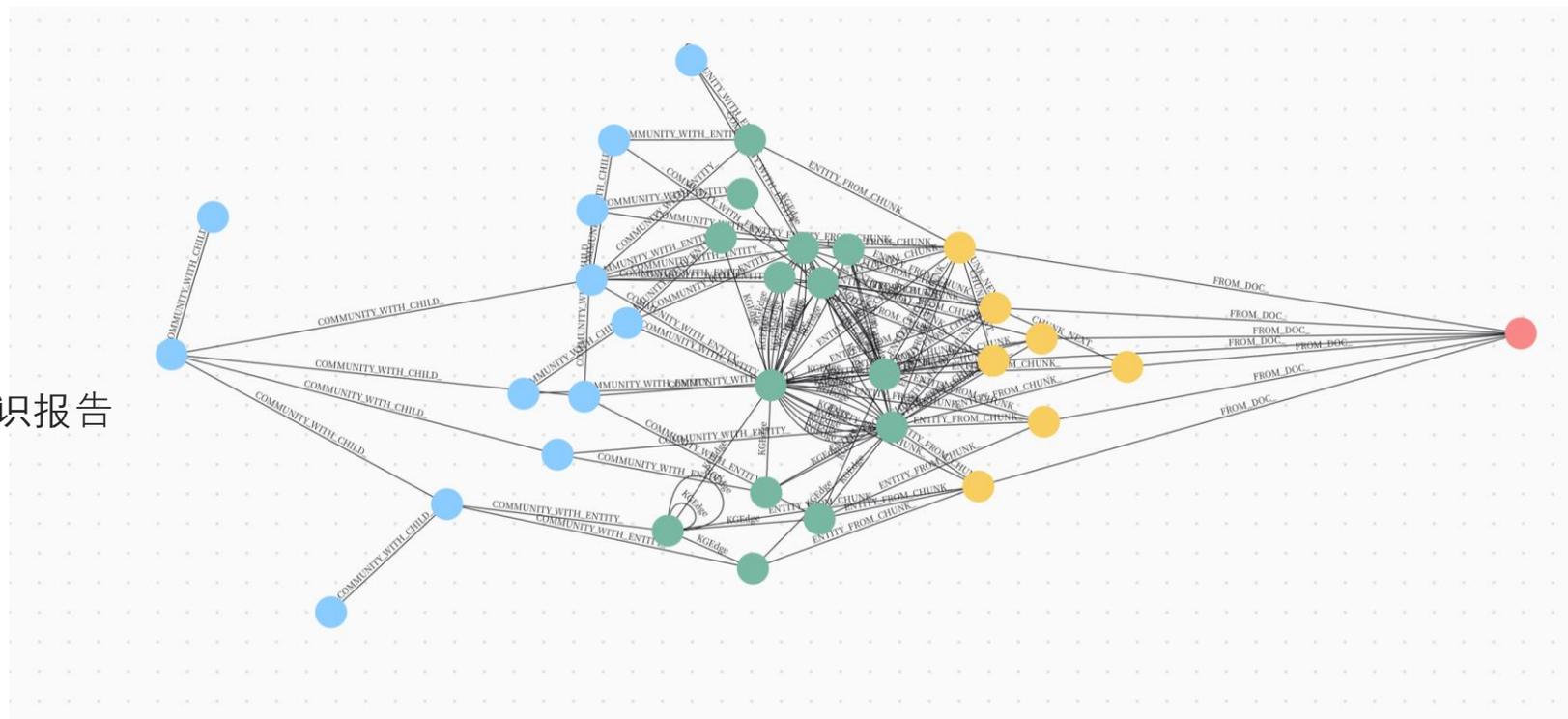
知识型：

注重总结查询结果的能力，使用自然语言总结回答



图谱的分层管理

按需可以在 NebulaGraph 上统一存储
元数据、原始数据、索引、知识图谱、知识报告
从而实现有计划的增删查改





📦 NebulaGraph Database Helper

在 NebulaGraph 中，如何引入全文索引的支持？

☰ Sources

NebulaGraphRAG 应用平台 和开发者 SDK

- 目前仅支持 7.x 版本的 Elasticsearch。

2. 配置 Elasticsearch:

- 为了支持外网访问 Elasticsearch，需要将 `config/elasticsearch.yml` 中的 `network.host` 设置为 `0.0.0.0`。
- 用户可以根据业务需求配置 Elasticsearch，如果需要定制 Elasticsearch，可以参考 [Elasticsearch 官方文档](#)。



📦 NebulaGraph Database Helper

在 NebulaGraph 中，如何引入全文索引的支持？

☰ Sources

NebulaGraph Chinese Docs
source • 1

nebulagraph chinese blogs
source • 2

🔗 Answer

在 NebulaGraph 中引入全文索引的支持可以按照以下步骤进行：

1. 部署 Elasticsearch 集群：

- NebulaGraph 的全文索引是基于 Elasticsearch 实现的，因此首先需要部署 Elasticsearch 集群。可以参考 [Kubernetes 安装 Elasticsearch](#) 或 [单机安装 Elasticsearch](#)。
- 目前仅支持 7.x 版本的 Elasticsearch。

2. 配置 Elasticsearch：

- 为了支持外网访问 Elasticsearch，需要将 `config/elasticsearch.yml` 中的 `network.host` 设置为 `0.0.0.0`。
- 用户可以根据业务需求配置 Elasticsearch，如果需要定制 Elasticsearch，可以参考 [Elasticsearch 官方文档](#)。

Present by

余毅航 (盐粒/Yanli)



> 在图上引入 LLM 的力量

For me: yanli.one



For NebulaGraph:



扫一扫上面的二维码图案，加我微信