



NebulaGraph x Airwallex

图数据库与风控

NebulaGraph Community Meetup · 上海站

 2025 年 3 月 29 日 上午 10:00

 上海市黄浦区黄陂南路 838号 中海国际 A 座 21 楼 Universe 会议室



Tim Zhu (朱天石)

Engineering Director at Airwallex

开场

嘉宾介绍

专注于支付和商户入驻相关的风险管控、技术战略与工程管理。Tim 负责构建高可用、可扩展的金融基础设施，保障全球支付网络的稳定运行，在金融科技领域的技术创新和团队管理经验，为 Airwallex 的全球化业务拓展提供了坚实的技术保障。

NebulaGraph 在 Airwallex 跨境金融风控的应用

郝鑫

Airwallex 风控平台 Engineer

NebulaGo Maintainer

xin.hao@airwallex.com

空中云汇的业务简介

全球企业账户

一站式开启多币种账户、高效国际付款、云汇 Visa 卡、费用管理和全球收单等服务。

全球收单

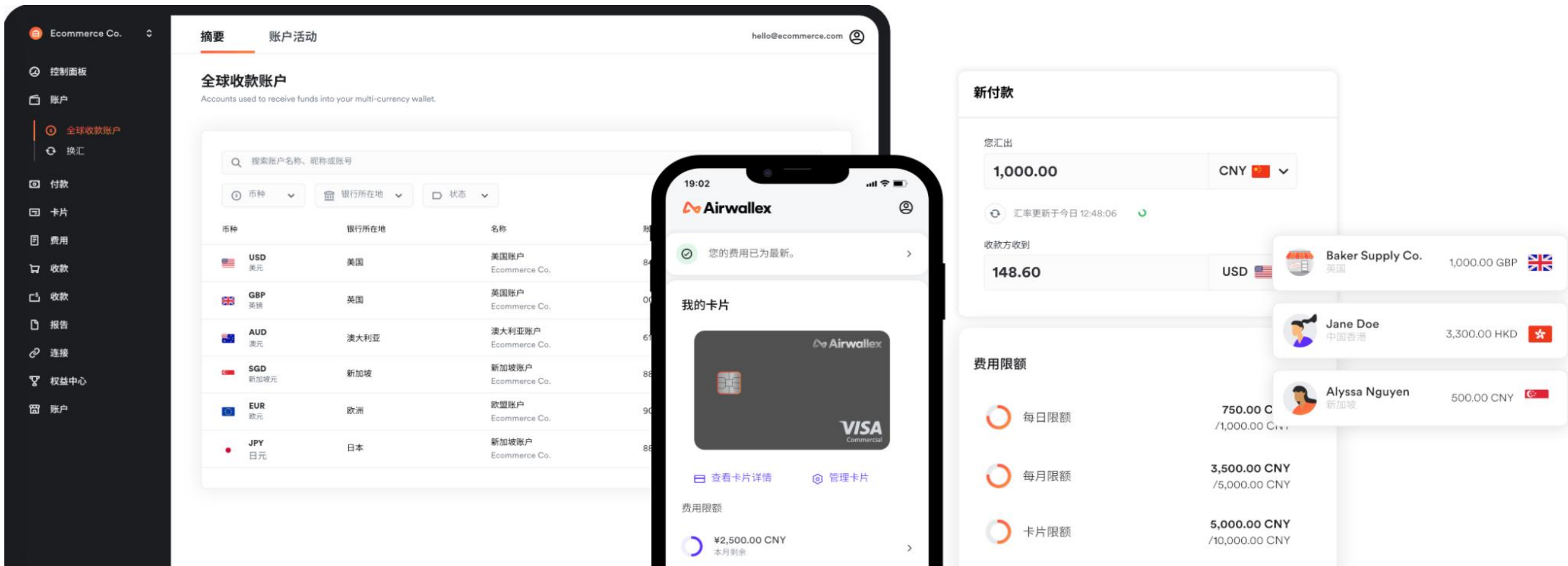
通过 Airwallex 的全球收单方案，企业可以轻松接受全球多个币种、多种支付方式的在线付款，全面提高支付成功率，同时节省换汇成本。

嵌入式金融

为平台提供嵌入式的金融服务，延展产品能力、优化您的用户体验，激发新的增长点。

全球支出管理

通过 Airwallex 空中云汇全球支出管理平台，节省财务运营时间，更清晰地掌控企业全球支出情况。



NebulaGraph 在空中云汇的使用场景

客户关系洞察

- 商业实体关联
- 个人信息关联
- 设备信息关联
- 网络信息关联
- 人脸信息关联

实时交易关联

- 设备信息关联
- 支付（收款）
人信息关联
- 账单信息关联
- 物流信息关联

模型特征提取

- 数据快照，用于离线训练
- 实时接口，计算特征，用于模型推断

可视化交互

- Web 页面，运营人员的可视化查询

客户关系洞察 (KYC) 的核心组件

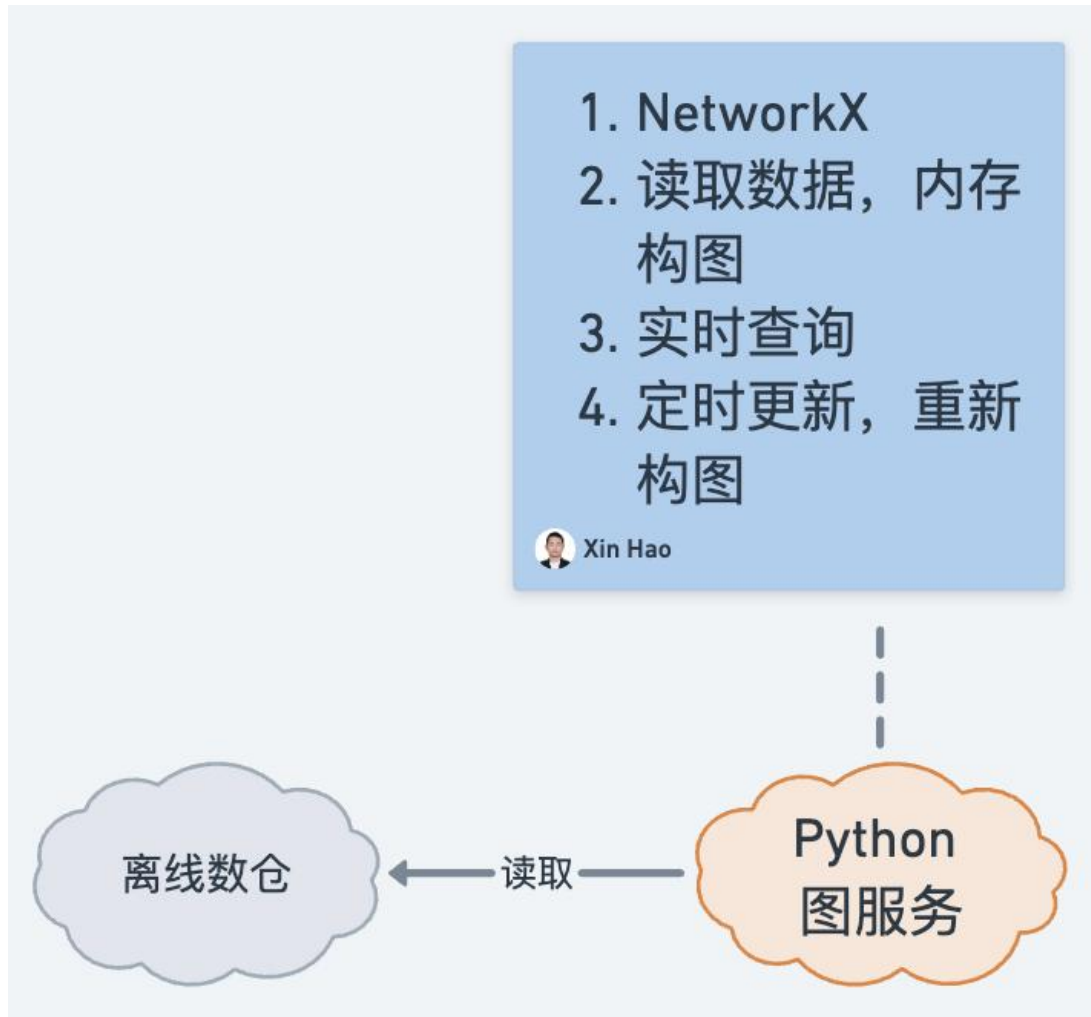
什么是 KYC?

KYC (know your customer) 是建立客户身份的过程，包括验证新客户的身分，以确保我们对客户、他们的风险状况和财务状况有足够的信息。它被认为是客户尽职调查流程的重要组成部分，以防止欺诈和金融犯罪。

图关系的引入

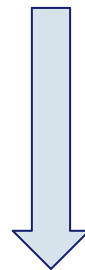
图关系是 KYC 需求的自然表征。

引入 NebulaGraph 之前



面临的问题

- 稳定性, 内存效率
- 表现力, 查询语言
- 迭代慢, 开发效率
-



引入图数据库

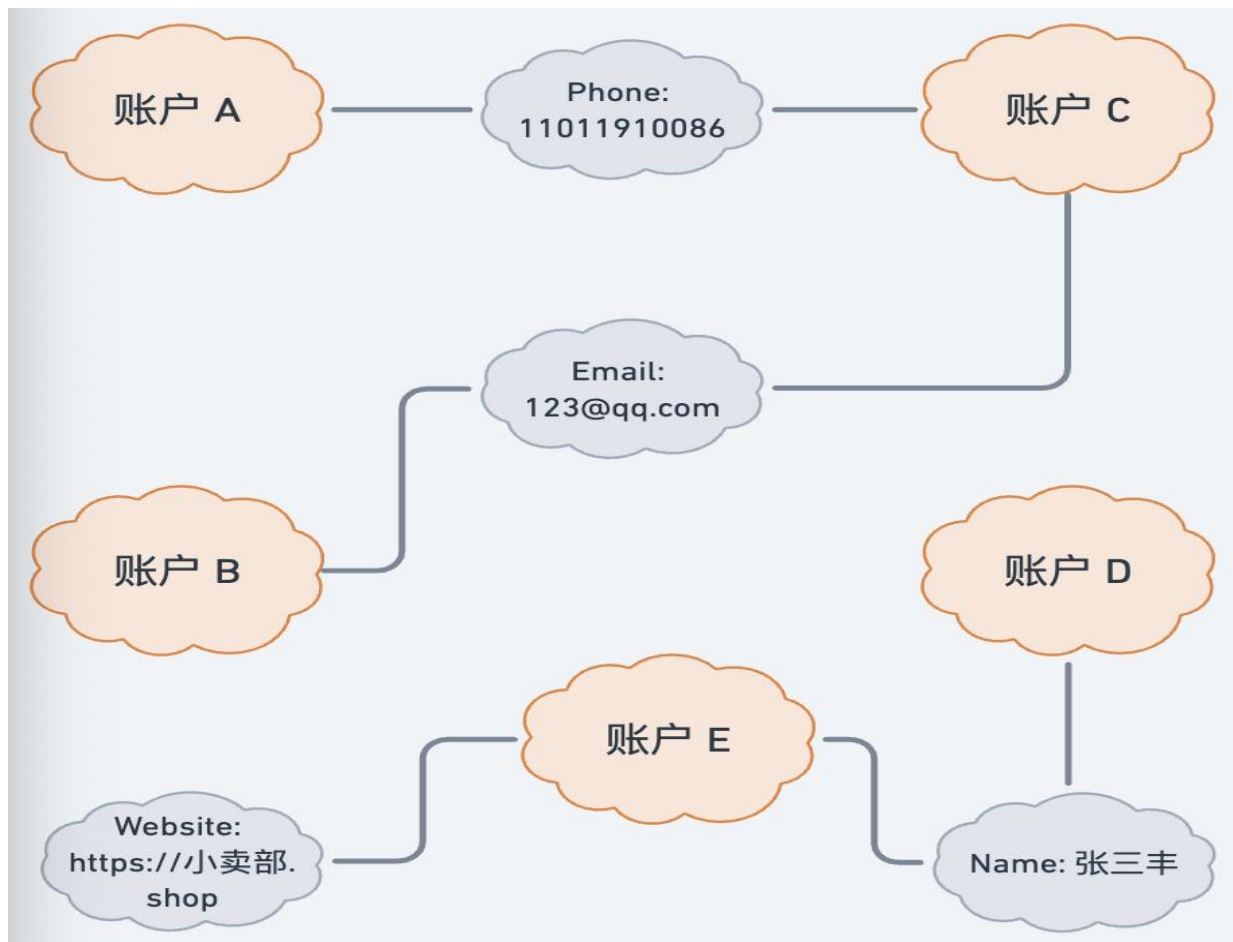
为什么选择 NebulaGraph?

- 功能全面，OLTP，OLAP，及开箱即用的基本算法支持
- 性能测试较为优异
- 查询语言满足需求
- 客户案例，企业支持

客户关系的构图

异构图

- 易于满足不同的查询场景
- 易于引入新的关系类型
- 易于结合 TTL 进行数据清理



客户关系系统架构

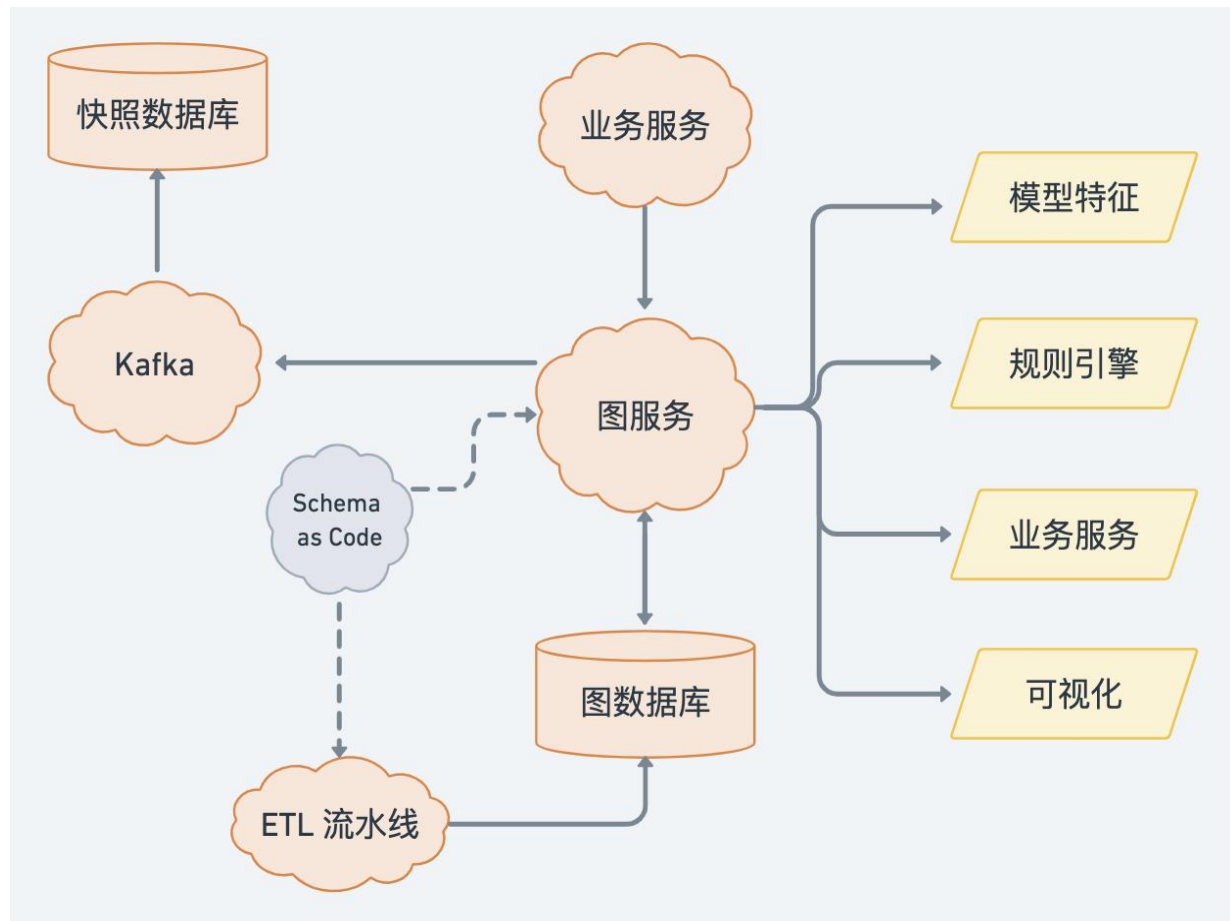
数据写入

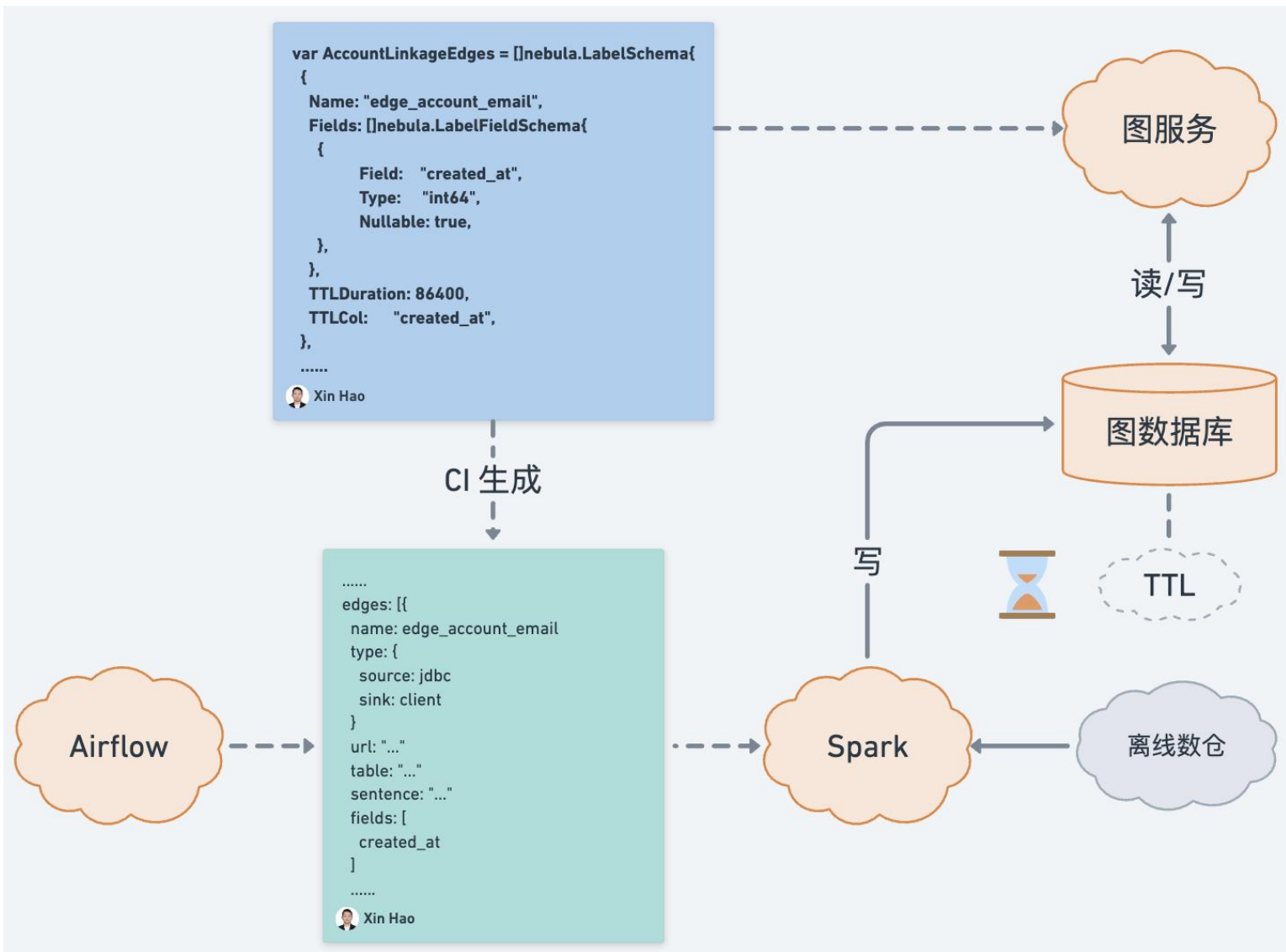
- 小时级，批处理 (全量)
- 业务系统，实时数据补齐 (增量)

Schema First

- Schema as Code
- 单一 Schema，既用于服务代码，同时生成 ETL 的 Schema

模型特征留存





To B 场景的业务复杂性

- Schema 复杂且易变
- 数据逻辑变更与清洗

如何应对?

- 一处定义，动态生成
- TTL + 全量写入

Schema: Declarative, not Versioned

```

- Field: "is_vpn",
+ Field: "is_anonymous",
  Type: "bool",
  Nullable: true,
},
+ {
+   Field: "city",
+   Type: "string",
+   Nullable: true,
+ },
+ {
+   Field: "created_at",
+   Type: "int64",
+   Nullable: true,
+ },
},
- TTLDuration: 86400,
- TTLCol: "created_at", // 86400 secs = 1 day
+ TTLDuration: 14400,
+ TTLCol: "created_at", // 14400 secs = 4 hours

```

ApplyEdge: add the not existing fields:
is_anonymous, city.

ALTER EDGE acc_ip ADD (is_anonymous
bool);

ALTER EDGE acc_ip ADD (city string);

ApplyEdge: remove the not expected
fields: is_vpn.

ALTER EDGE acc_ip DROP (is_vpn);

ApplyEdge: alter the edge TTL, duration
from 86400 to 14400.

.....

查询结果映射

声明即所得

```
type SubGraphRow struct {
→ Nodes []*Node `nebula:"nodes"`
→ Edges []*Relationship `nebula:"relationships"`
}

type Node struct {
→ Vid string `nebula:"_vid"`
→ TagName string `nebula:"_tag_name"`
→ BillingEmail string `nebula:"billing_email"`
→ // .....
}

type Relationship struct {
→ Src string `nebula:"_src"`
→ Dst string `nebula:"_dst"`
→ Email string `nebula:"email"`
→ // .....
}
```

```
rs, err := NebulaClient.ExecQuery("...")
if err != nil {
→ return nil, err
}

var rows []*schema.SubGraphRow
err = rs.Scan(&rows)
if err != nil {
```

跨境交易风控系统的组件

在空中云汇，实时风控引擎为我们的客户提供了一种解决方案，以打击其在线支付系统中的欺诈活动。通过利用机器学习模型和实时数据分析，我们的风控引擎帮助企业减少欺诈和退款，同时确保为客户提供安全无阻的支付体验。

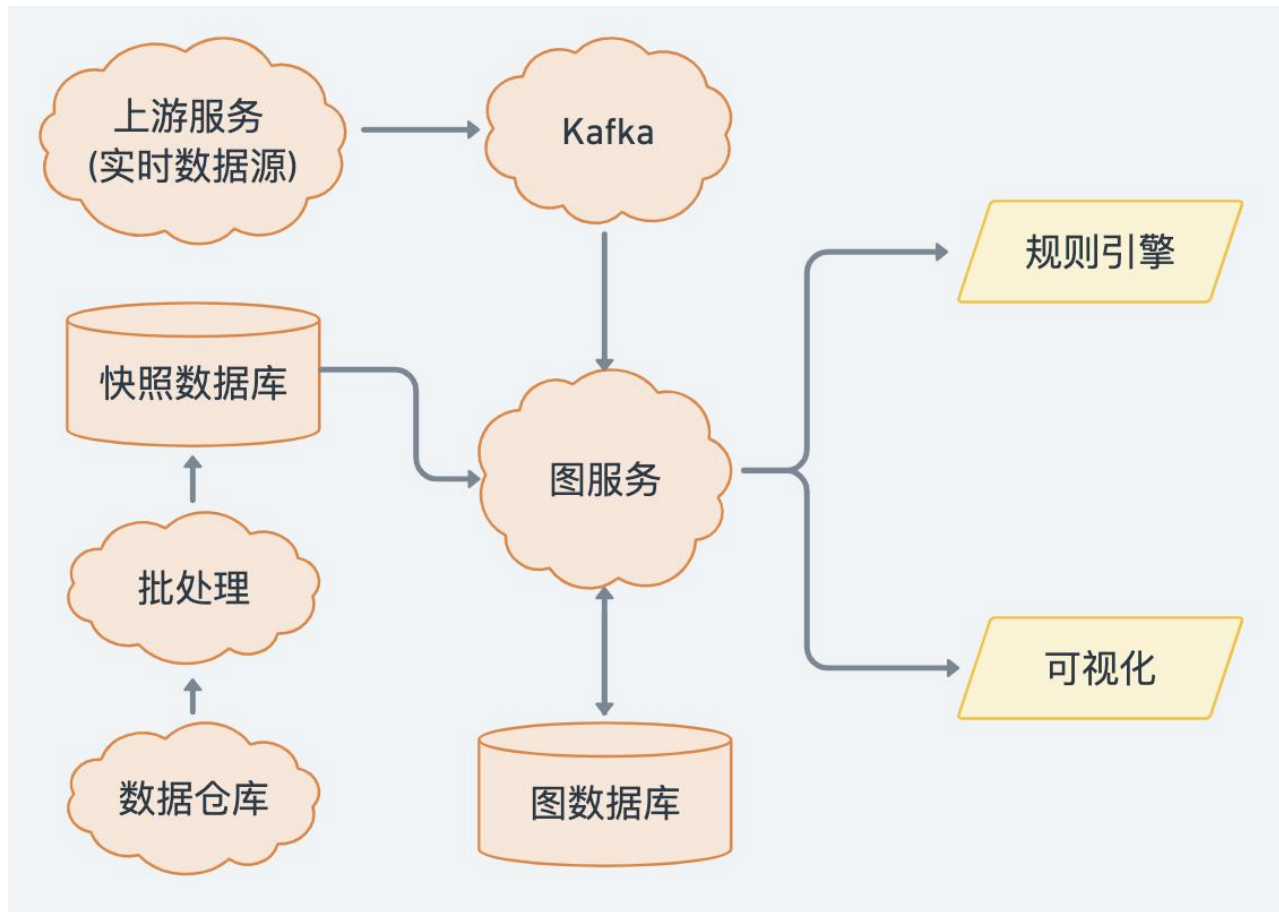
跨境支付与境内支付的差异

- 跨境支付对于交易的延迟，容忍度比境内支付要高：一般数秒。所以留给实时风控的时间缓冲较多一些。
- 跨境支付能够获取的支付者信息以及支付场景信息相较境内支付要少。
- 跨境支付的方式相较境内支付更传统，风控难度更高。

交易关联系统架构图

分离计算，读时合并

- 历史数据 (> 4h) 小时级批处理
- 实时数据 (< 4h) 图数据库查询
- 历史数据: 预计算子图点边; 内存计算图特征.



未来的展望

- 目前，单一 Nebula 集群，不同 Space 的计算资源不隔离，带来两个问题：
 - a. 不同应用间查询峰值会互相干扰；
 - b. 技术选型时，一套集群同时满足 TP 与 AP 的目标未实现。
- 图算法的支持：
 - a. 开放性不足；
 - b. 没有形成较丰富的生态。

谢谢 Q&A



<https://nebula-graph.com.cn>



GitHub: [vesoft-inc/nebula](https://github.com/vesoft-inc/nebula)



Twitter: [@NebulaGraph](https://twitter.com/NebulaGraph)



Facebook: [@NebulaGraph](https://www.facebook.com/NebulaGraph)



<https://discuss.nebula-graph.com.cn>



微信公众号



开源项目

NebulaGraph 在震坤行的应用和探索

分享人：黄春晖

目录

01

引言

02

NebulaGraph 在震坤行的应用实践

03

NebulaGraph 在震坤行的持续探索

04

总结与展望



PART. 01

引言

震坤行工业超市是一家 数字化的工业用品服务平台

帮助客户实现透明、高效、降成本

协同合作伙伴持续优化供应能力

为客户提供质量可靠、价格更优的工业用品

利用数字化的管理工具和物联网技术

帮助合作伙伴实现从采购、制造到交付的供应链全链路的数字化转型

自建仓储物流体系

致力于为合作伙伴提供覆盖全国的更适合工业用品的、成本更低的仓储、物流及现场交付服务

1998 做经销 震坤行成立

获得多家国际知名品牌胶
粘剂、润滑剂的代理权。

2008 做产品 深圳轴心自控

深圳轴心成立，提供全套点胶
技术解决方案。2016年被国际
龙头Mycronic收购。

2014 做电商 震坤行工业超市

正式更名为“震坤行工业超市
(上海)有限公司”，以胶粘
剂和特种润滑客户为种子客户，
成功发力MRO业务。

2023 成功上市 纽交所上市

2023年12月15日，正式
在纽约证券交易所挂牌上
市，股票代码为“ZKH”。

至今 做平台 快速发展

建成 **32个总仓、96个服务中心**
仓

100,000⁺家累计合作客户

4,000⁺名员工全力以赴

1,000⁺专业的销售与技术支持团队

200⁺华南专业的产销与技术团队

发展布局：数字化的工业用品服务平台

核心交易平台

大客户

ZKH 震坤行

小客户

邦 工邦邦

旗下品牌

产品开发

ndanda
安丹达

ZKH

物料管理

KUNTON

仓储物流

KUNHE 坤合

基础设施

- ① **大数据**（商品数据、客户需求数据、库存数据、设备使用及维护数据）
- ② **IT系统**（官网&APP、客户专属商城、采购系统智联、震坤行商家开放平台、智能小仓库）
- ③ **服务网络**（仓储物流、服务中心、技术服务、咨询服务）

32 条产线 | 1,700 万⁺SKU

备品备件、通用耗材、行政物资、加工制造、化学品

润滑剂	胶黏剂	建材	涂料记号标识	工厂自动化	通用设备	个人防护	安防
车间化学品	化学试剂	仪器仪表	电线电缆	清洁用品	五金&工具	焊接	搬运存储
电气自动化	低压配电	泵阀管件	气动液压	工装服饰	胶带标签	实验室仪器耗材	办公用品
动力传动	紧固密封件	磨具磨料	刀具量具	企业福利	照明	制冷暖通	家具&装修材料



累计合作客户
100,000+



震坤行的业务挑战

01



数据规模与复杂性的双重压力

业务扩展导致客户-供应商关系、客户行为等复杂关系数据快速增长，传统数据库难以高效存储和查询，实时分析需求难以满足。

02



关系数据库的局限性

关系数据库在多表连接等操作上性能瓶颈显著，无法支持复杂关系数据的实时分析和深度挖掘，影响供应链优化等高价值场景落地。

03



数据价值落地困难

复杂关系数据在供应链优化、风险预警等场景中的应用，面临数据建模、算法优化和效果验证等挑战，技术与业务深度融合是关键。

选择 NebulaGraph 的原因

高效的关系表达

通过图结构直观表示实体及其关系，节点表示实体（如**客户、供应商、商品**），边表示关系（如**供应、购买**）。特别适合震坤行复杂的供应链网络，能清晰呈现多层次的**客户-商品-供应商**关系

强大的路径分析能力

支持高效的图遍历算法，能够快速进行路径分析（如**最短路径、共同邻居**等），为震坤行提供强大工具，用于挖掘客户需求、优化供应链路径和识别客户风险

实时查询与响应

具备高并发、低延迟的查询能力，支持实时推荐、风险预警等场景，满足震坤行对实时性的需求，助力用户行为分析和供应链动态调整，快速响应业务变化

可扩展性与高性能

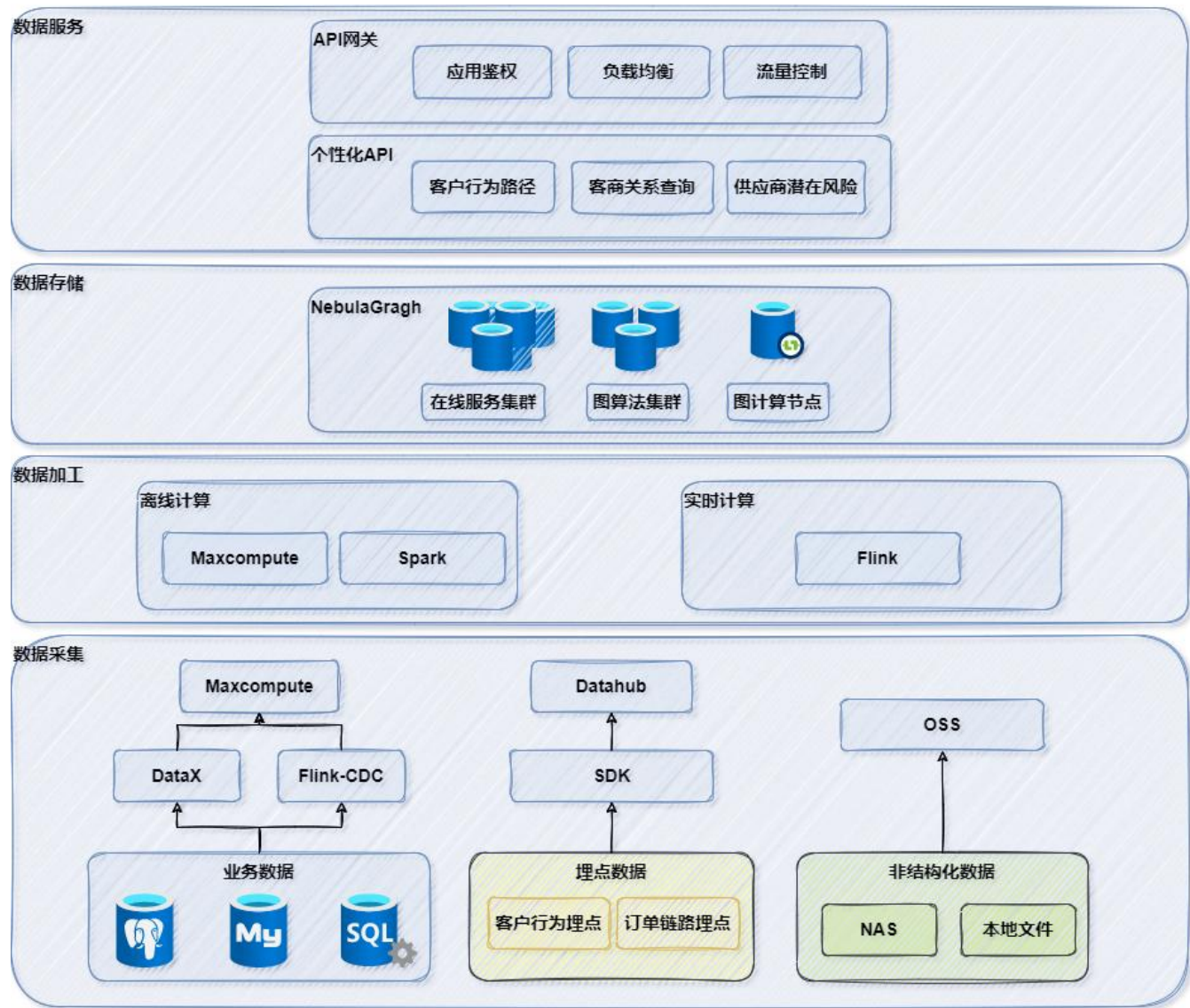
采用分布式和存算分离的架构，支持水平扩展，轻松应对震坤行数据量的快速增长，确保高并发场景下依然保持高性能



PART. 02

NebulaGraph 在震坤行的应用实践

NebulaGraph 在震坤行的技术架构



工业品电商的精准推荐之痛

01

依赖关键词匹配

主要基于关键词匹配，缺少用户的上下文数据，导致对用户意图的理解较为表面

02

个性化能力不足

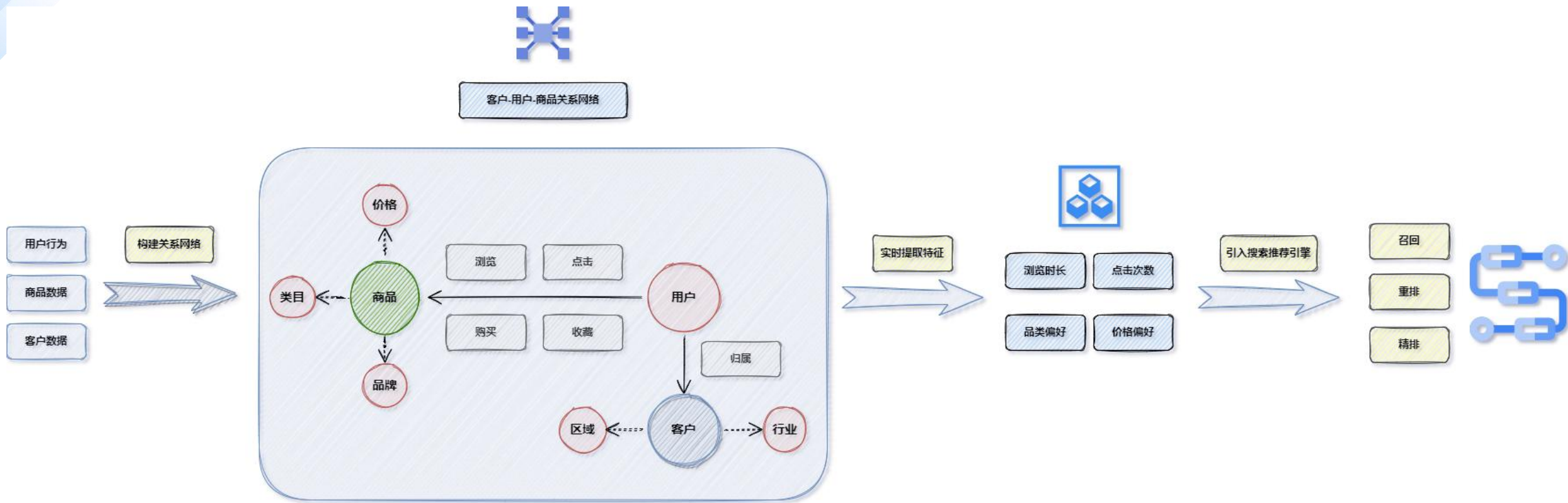
缺乏用户-商品关系数据支持，难以深入挖掘用户行为与商品之间的深度关联（如相似推荐、偏好分析），在提供高度个性化推荐内容方面存在明显不足

03

新客冷启动问题

新用户缺乏历史数据，难以生成有效的个性化推荐，搜索结果更偏向通用商品

数据驱动与算法优化



数据驱动与算法优化



01

构建用户-客户-商品关系网络

利用 NebulaGraph 高效存储和处理客户-用户-商品关系数据，构建多层次的关系网络，并且支持实时数据更新。



02

提升个性化推荐能力

算法召回和排序阶段引入用户行为特征（浏览时长、点击次数）和商品偏好（用户对品类、品牌、价格的偏好）特征，提升推荐精准性和多样性



03

新客智能化冷启动

利用大模型识别新用户的行业及地址信息，针对性的推荐同行业、同区域客户的常买清单商品

搜索推荐优化-成果总结与未来规划

01

用户体验与业务增长双赢

- 通过引入图数据库与智能化算法，成功优化搜索推荐系统，深度挖掘用户-商品关系，满足用户个性化需求。
- 搜索推荐点击率提升**15%**
- 下单转化率提升**5%**

02

持续优化与未来展望

- 拓展更多关系数据维度（如客户需求采集），进一步提升推荐精准性。
- 引入AI实时学习机制，动态适应用户行为变化，实现推荐系统的自我优化

供应链信息孤岛痛点

01.

采购成本增加

- 信息割裂：各供应商的价格、库存及交付能力信息无法整合，影响采购决策的全面性。
议价劣势：缺乏需求数据和历史交易记录，在谈判中难以争取有利条件，导致采购成本上升。
紧急采购多：因需求预测不准确或库存失衡，频繁进行高成本紧急采购，增加了额外支出。

03.

履约协调不足

- 仓库选择混乱：无统一库存视图，难选最优发货仓库，增加运输成本和交货不确定性。
订单准确性差：数据未共享，订单与拣选信息不对称，影响商品匹配和客户体验。
履约效率低：信息不连贯，订单处理流程效率低，影响整体履约效能。

02.

库存管理失衡

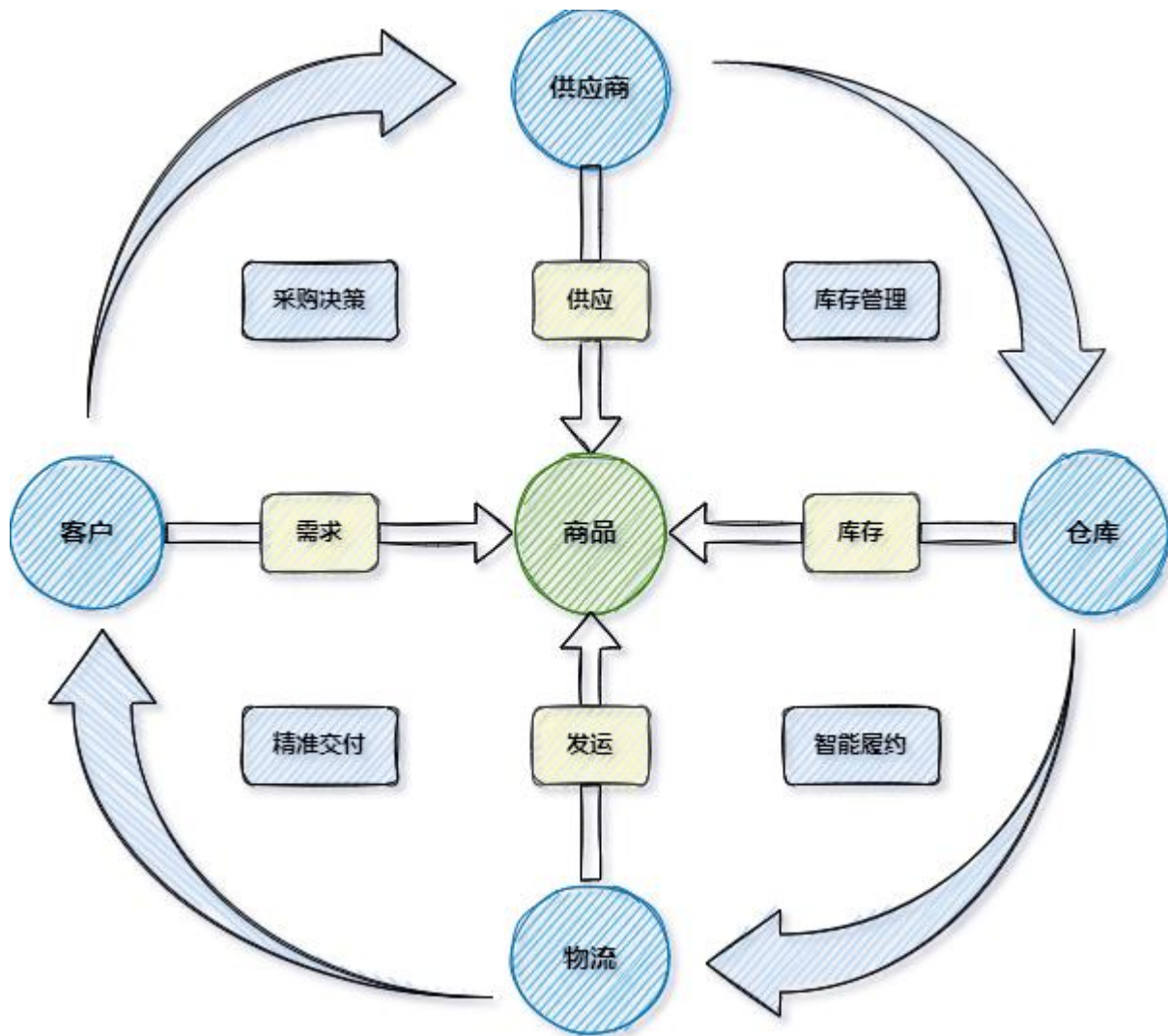
- 库存数据割裂：仓库与商品库存难关联，全局可视性差，影响决策。
补货效率低：系统未同步销售和库存数据，补货决策不准，不及时或过量。
仓储成本增加：商品库龄长，增加仓储空间需求及存储成本。

04.

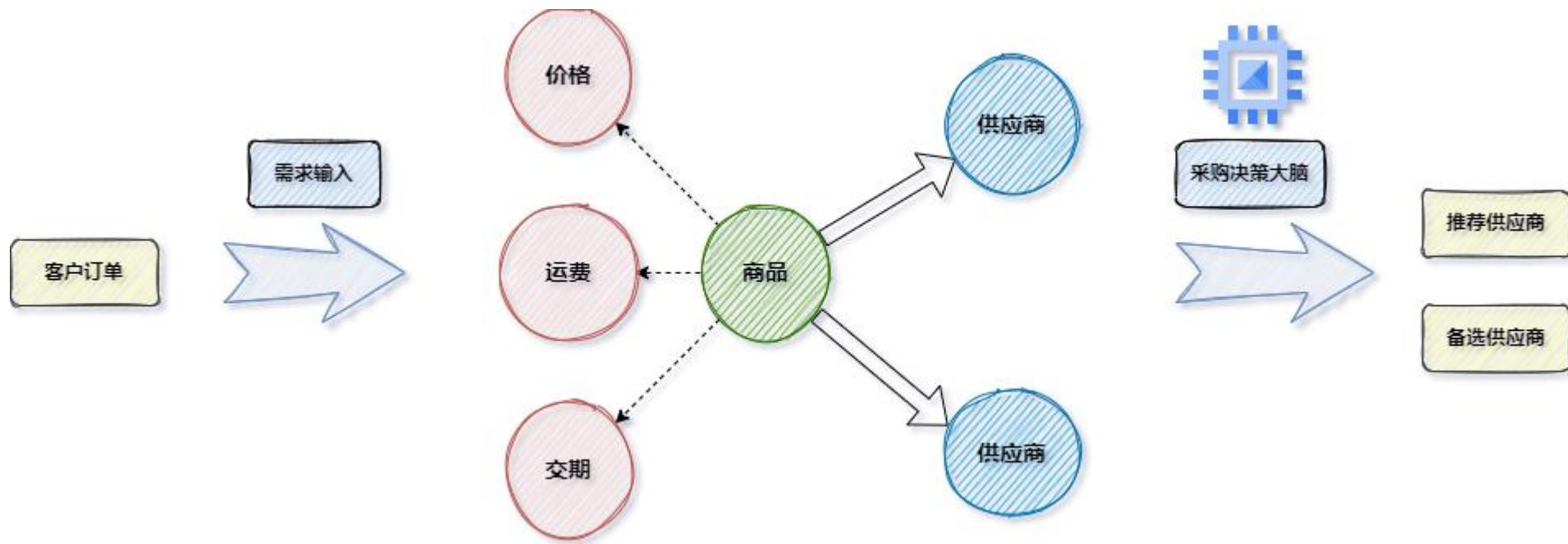
交付精准不足

- 信息不通畅：物流信息查询难，交货时间预测不准，增加不确定性和客户焦虑。
路径规划低效：缺乏数据支持，配送路线不佳，运输成本高且易延误。
服务体验受损：信息断层，交货不准时或错误，影响客户信任，可能导致退货和负面评价。

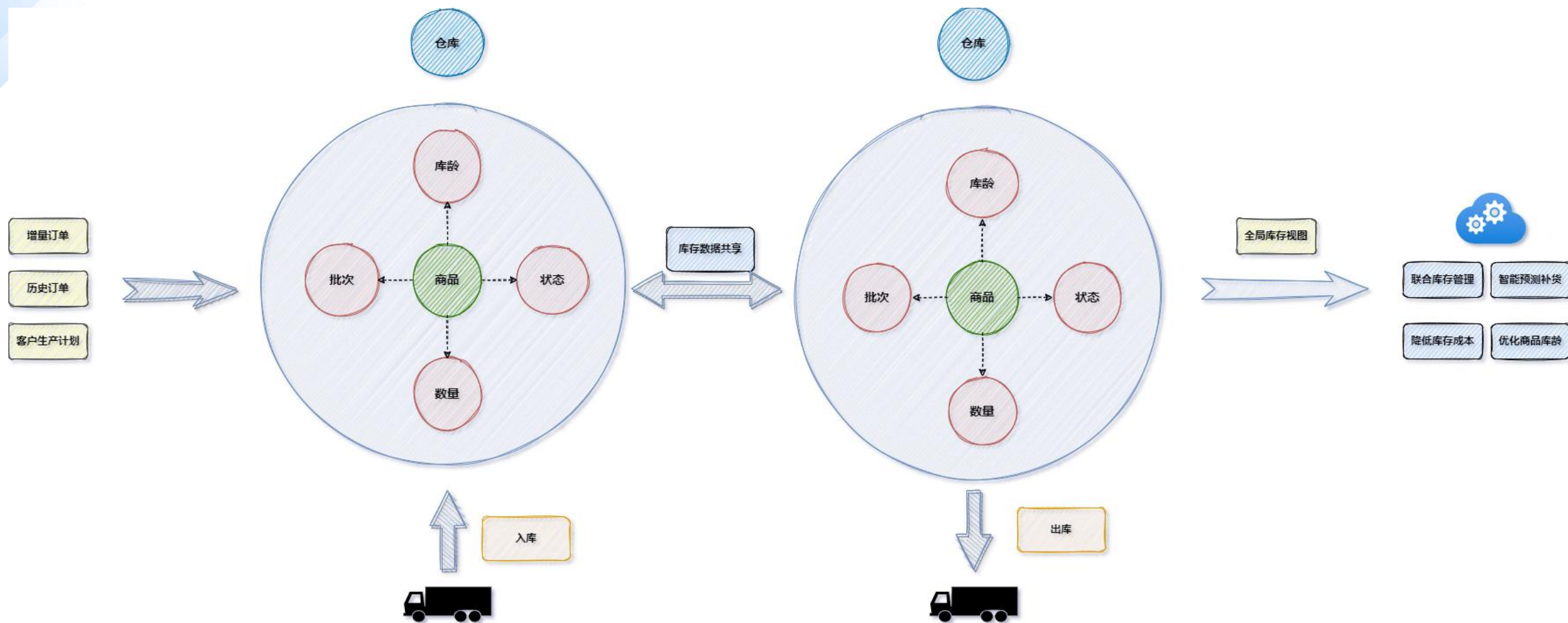
NebulaGraph 支持下的高效供应链网络构建



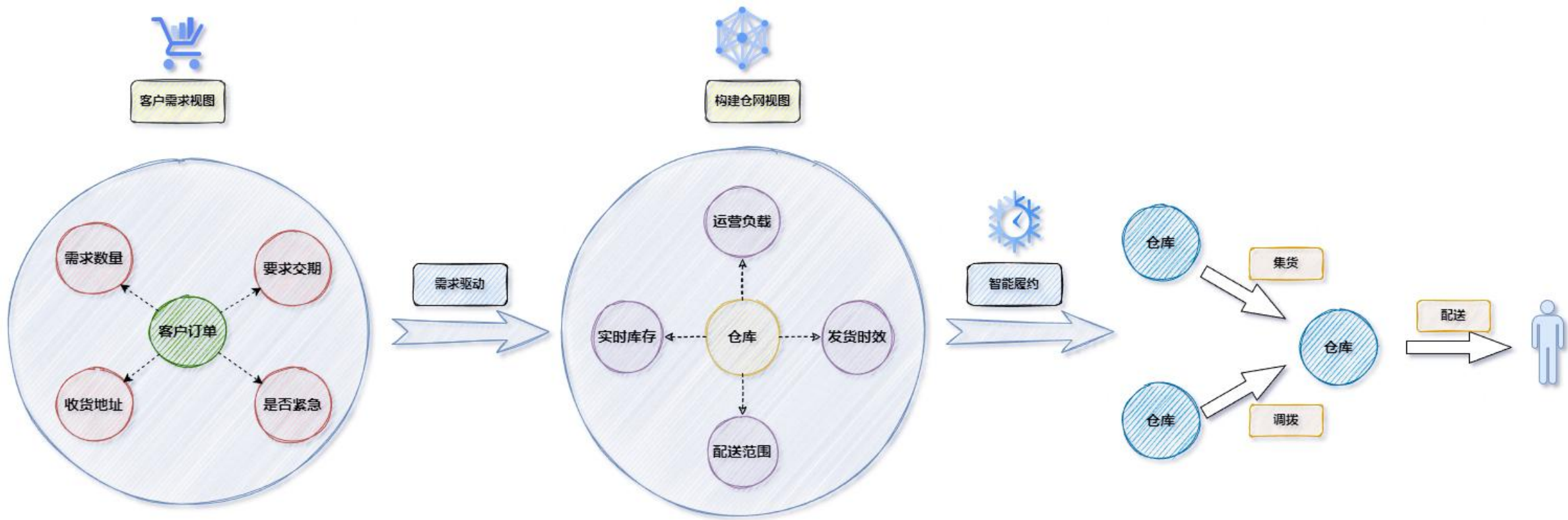
商品供应网络驱动采购决策



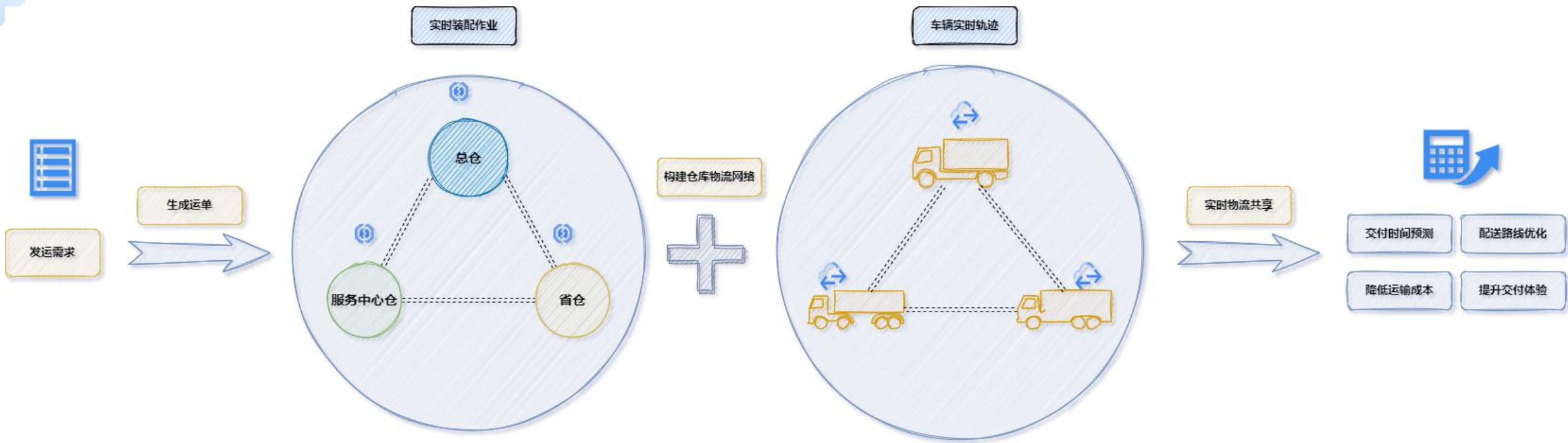
仓储全局网络优化库存管理



供应链网络协调智能履约



物流网络助力精准交付



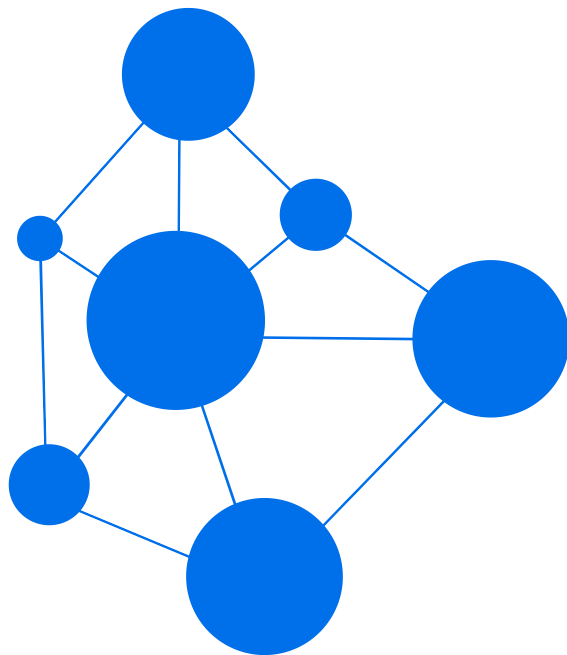
从信息孤岛到全局优化的智能化转型

客户满意度提升

- 透明的物流信息、精准的交付和高效的履约流程可以显著提升客户满意度。
通过数据驱动的服务优化，能够建立长期的客户信任和忠诚度。
- 客户售后工单数减少**20%**。

成本降低与效率提升

- 通过优化库存管理、采购策略和配送路线，可以显著降低运营成本。
数据驱动的精准决策能够提升整体运营效率，改善客户体验。
- 有货率提升**15%**，呆滞库存减少**10%**。



数据关联与实时分析

- 图数据库能够高效处理供应链中复杂的多维度关系，实现数据的全局可视化和实时分析。
通过数据驱动的决策，可以显著提升供应链的效率和稳定性。

智能化与自动化

- 基于NebulaGraph的图算法（如**最短路径算法**、**协同过滤算法**等）可以自动优化库存、采购、履约和配送等环节，减少人为干预。
智能化的供应链系统能够快速响应市场变化，提升竞争力。



PART. 03

NebulaGraph

在震坤行的持续探索

知识库管理现状

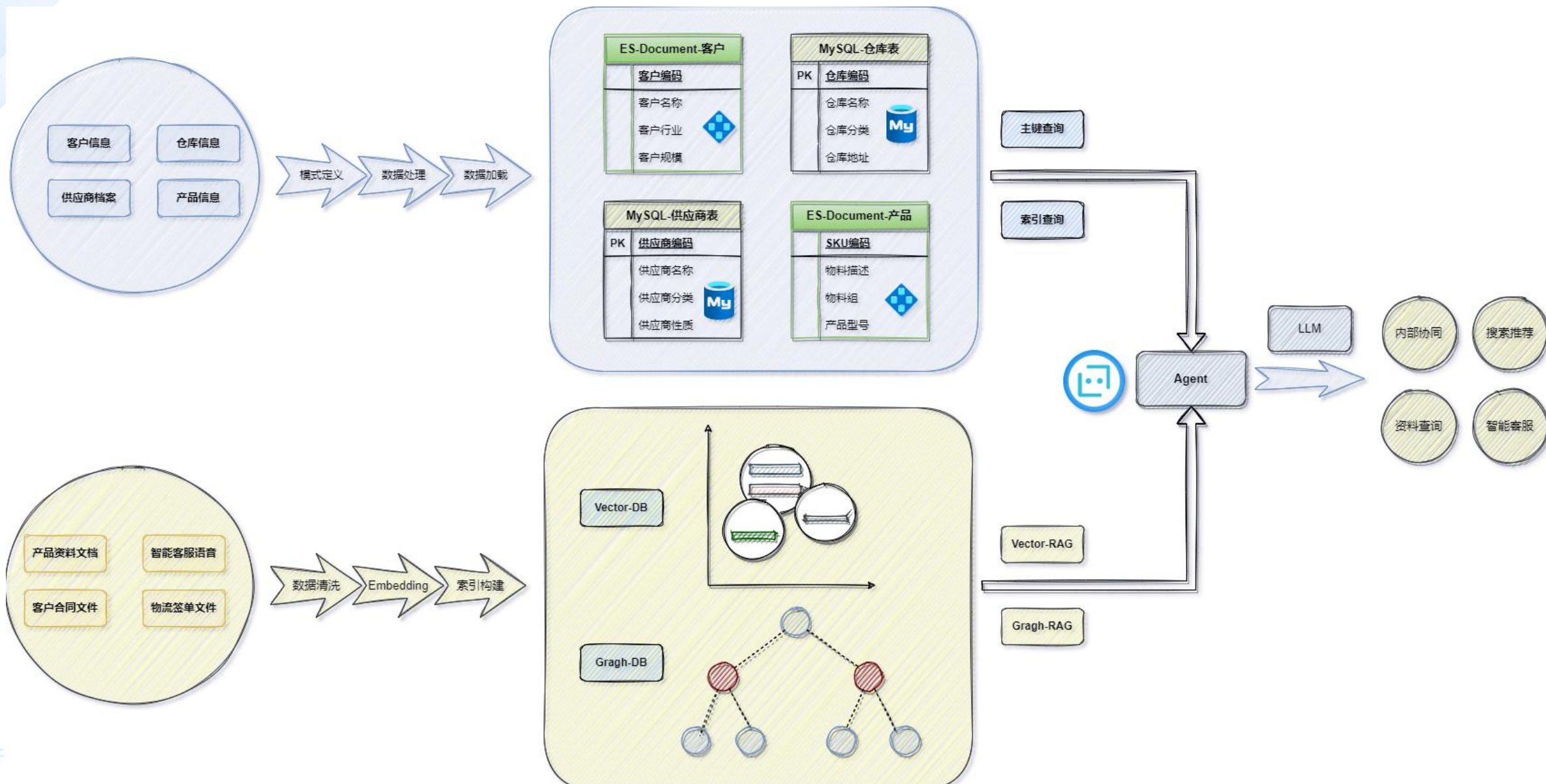
结构化数据为主，非结构化数据管理不足

- 存储与查询高效：当前知识库以结构化数据为主，使用关系数据库（如MySQL）或 Elasticsearch存储，支持高效查询和统计分析。
覆盖范围有限：结构化数据难以满足复杂、多样化的知识需求，无法有效管理非结构化内容。

非结构化数据管理不足，价值未能充分释放

- 管理分散且检索困难：非结构化数据（如产品资料、客户合同文件等）分散存储，缺乏统一管理平台，难以快速检索和利用。
业务赋能不足：非结构化数据的价值未被充分挖掘，未能深度融入业务活动（如智能客服、搜索推荐等），限制了知识库的整体效能。

知识库架构规划



工业品电商信用风控挑战

01

信息不对称，风险评估难度大

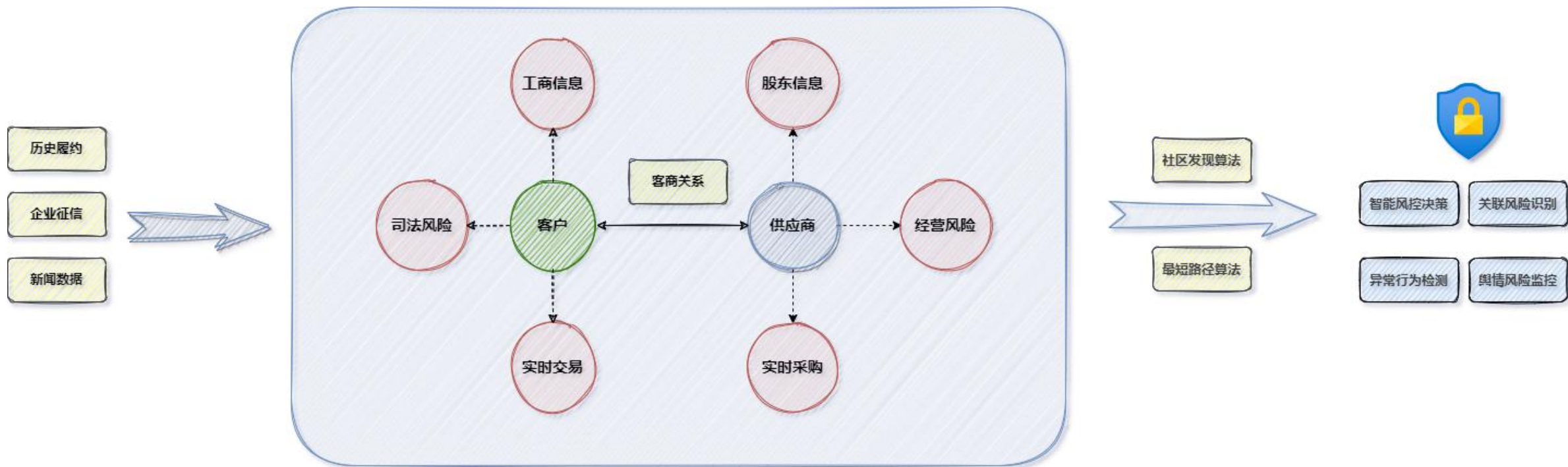
风控系统依赖规则和人工审核，难以全面获取和分析多源数据，导致信息不对称。风险评估缺乏全面性和准确性，难以识别潜在风险，增加了业务运营的不确定性。

02

欺诈手法复杂多变，识别难度加大

欺诈手段采用高级技术和社交工程，传统风控方法难以有效应对。企业面临更高的财务损失和声誉受损风险，需要更智能的策略来识别和防范不断变化的威胁。

NebulaGraph 助力构建智能风控体系





PART. 04

总结与展望

NebulaGraph 应用价值总结

01

提升业务效率

通过客户-商品-行为关系图实现精准推荐，提升点击率和转化率；通过供应商-商品-物流关系图优化供应链，提升库存周转并降低物流成本。

02

降低运营成本

图数据库的高效关联查询和实时分析能力，减少了传统数据库的复杂查询开销，降低系统维护成本；通过优化供应链路径，减少资源浪费，节省运营成本。

03

优化决策支持

提供路径分析和关系挖掘能力，支持精准决策；通过可视化展示复杂关系网络，帮助管理层直观理解业务动态。

04

增强业务创新能力

图数据库推动业务模式创新，如在知识库管理和信用风控等领域的探索，为智能化转型奠定基础。

未来合作展望



01

探索新应用场景

构建客户-问题-解决方案关系图，实现智能客服；分析市场趋势和客户行为，支持市场预测；优化员工-技能-项目关系图，提升团队协作效率。



02

推动技术赋能业务

挖掘图数据库潜力，构建业务洞察体系，支持实时决策；结合 AI 技术，应用于智能风控和动态定价；通过行为分析和需求预测，优化资源配置与业务策略。



03

提升数据价值与业务协同

通过图数据库优化数据整合与共享，提升跨部门协同效率；挖掘数据潜在价值，支持业务创新与战略规划；构建数据驱动的业务闭环，推动企业数字化转型。

谢谢大家 Q&A



<https://nebula-graph.com.cn>



GitHub: [vesoft-inc/nebula](https://github.com/vesoft-inc/nebula)



Twitter: [@NebulaGraph](https://twitter.com/NebulaGraph)



Facebook: [@NebulaGraph](https://www.facebook.com/NebulaGraph)



<https://discuss.nebula-graph.com.cn>



微信公众号





开源项目

NebulaGraph x Airwallex

图数据库与风控

NebulaGraph Community Meetup · 上海站

 2025 年 3 月 29 日 上午 10:00

 上海市黄浦区黄陂南路 838号 中海国际 A 座 21 楼 Universe 会议室

茶歇时间
10 min 后精彩继续!



微信公众号



上海站活动群

业务增益32% NebulaGraph 在携程风控场景下的落地实践

周伟诚 携程风控团队

分享人

周伟诚

- 携程金融风险管理部 风控研发
- 南京大学硕士
- 特征平台研发工程师, NebulaGraph布道师
- 授权专利2项;全国水中机器人大赛单项冠军



目录

1

风控系统概述

2

NebulaGraph
在携程风控的落地实践

3

总结和展望

1

风控系统概述

风控概述



定义

在特定的风险环境中，
通过管理手段将业务风险降至最低。



目标

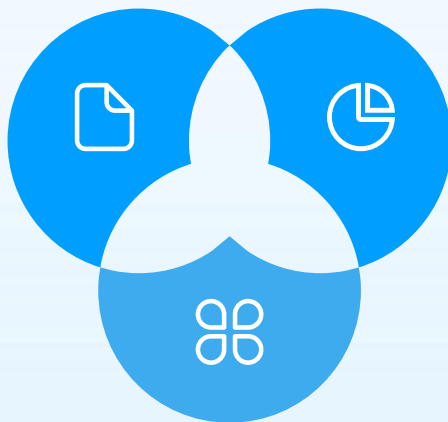
更小的交易损失量
更低的用户打扰率



风控三要素

规则策略

依赖数据进行决策



数据特征

实时数据
历史数据

人力支持

支持风控流程的专业团队



团队构成



策略分析

营销作弊, 支付欺诈, 消费金融

数据算法

策略评估, 模型建设

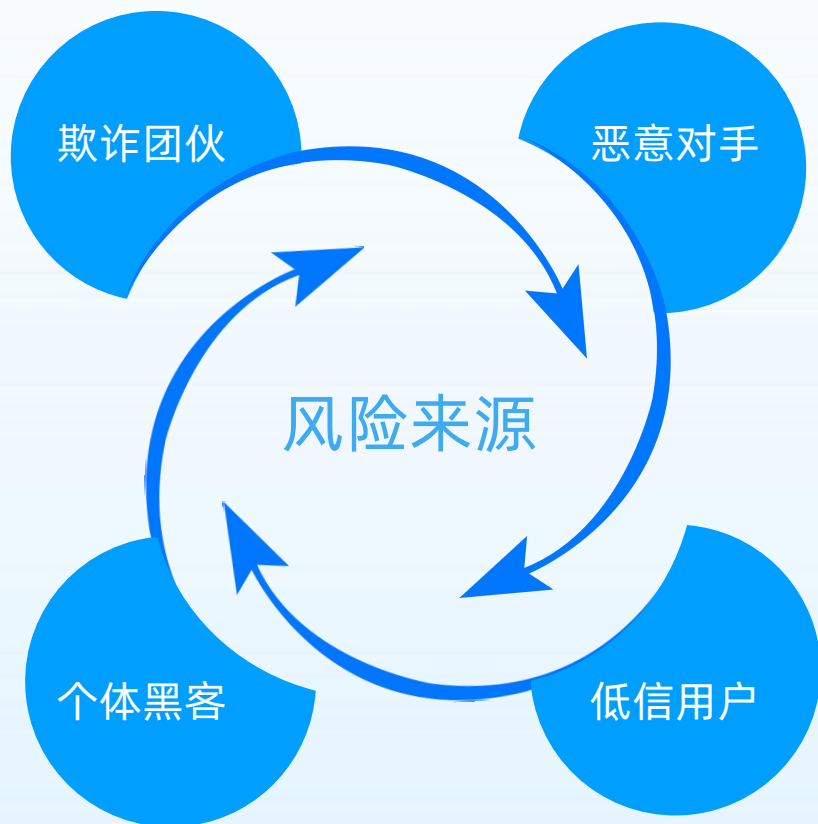
系统研发

平台开发, 数仓建设

运营审核

人工判定, 事后跟进

常见的风控场景



平台风险

酒店: 骗赔, 刷单, 占房
机票: 骗保, 私退, 虚占
市场: 刷点评, 优惠券



支付风险

盗卡, 洗钱, 内外卡拒付



消费金融风险

欺诈风险: 身份伪造

常见风控案例

01

平台风险:骗赔

用户/商户利用携程的赔付政策和服务流程. 夸大事实或虚构赔付场景, 骗取赔付.

02

支付风险:电信诈骗

诈骗分子通过社交平台物色未成年人, 诱骗未成年人配合操作购买机票.

03

消费金融:身份伪造

欺诈分子使用**假冒或伪造身份**的手段进行资金借贷, 拒绝还款

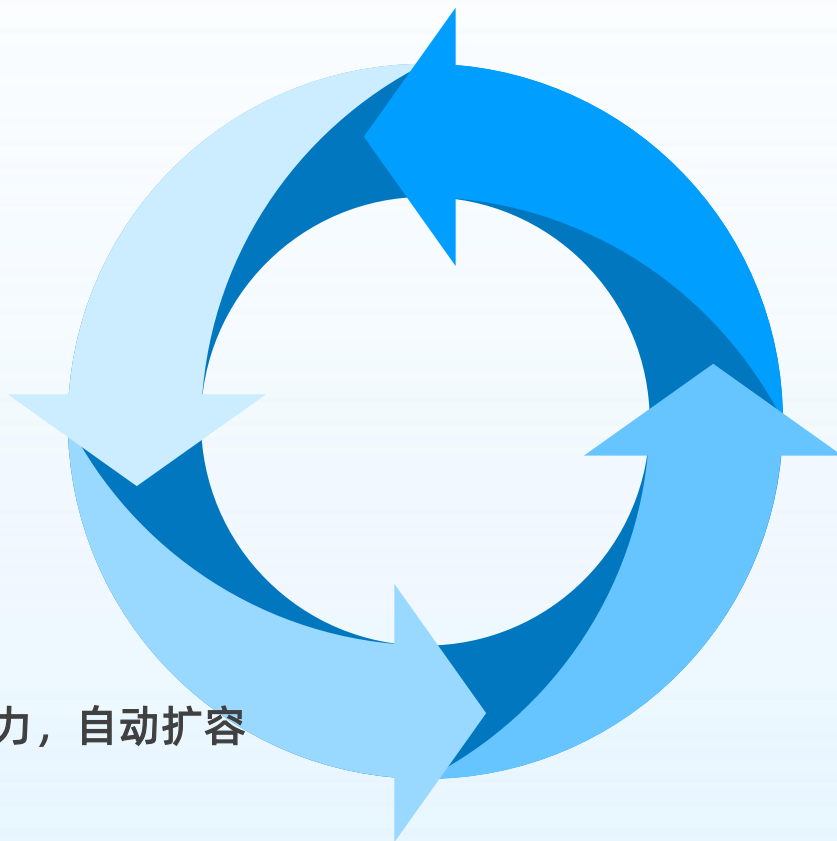
风控系统的保障

高吞吐量
每日做出几十亿次同步决策

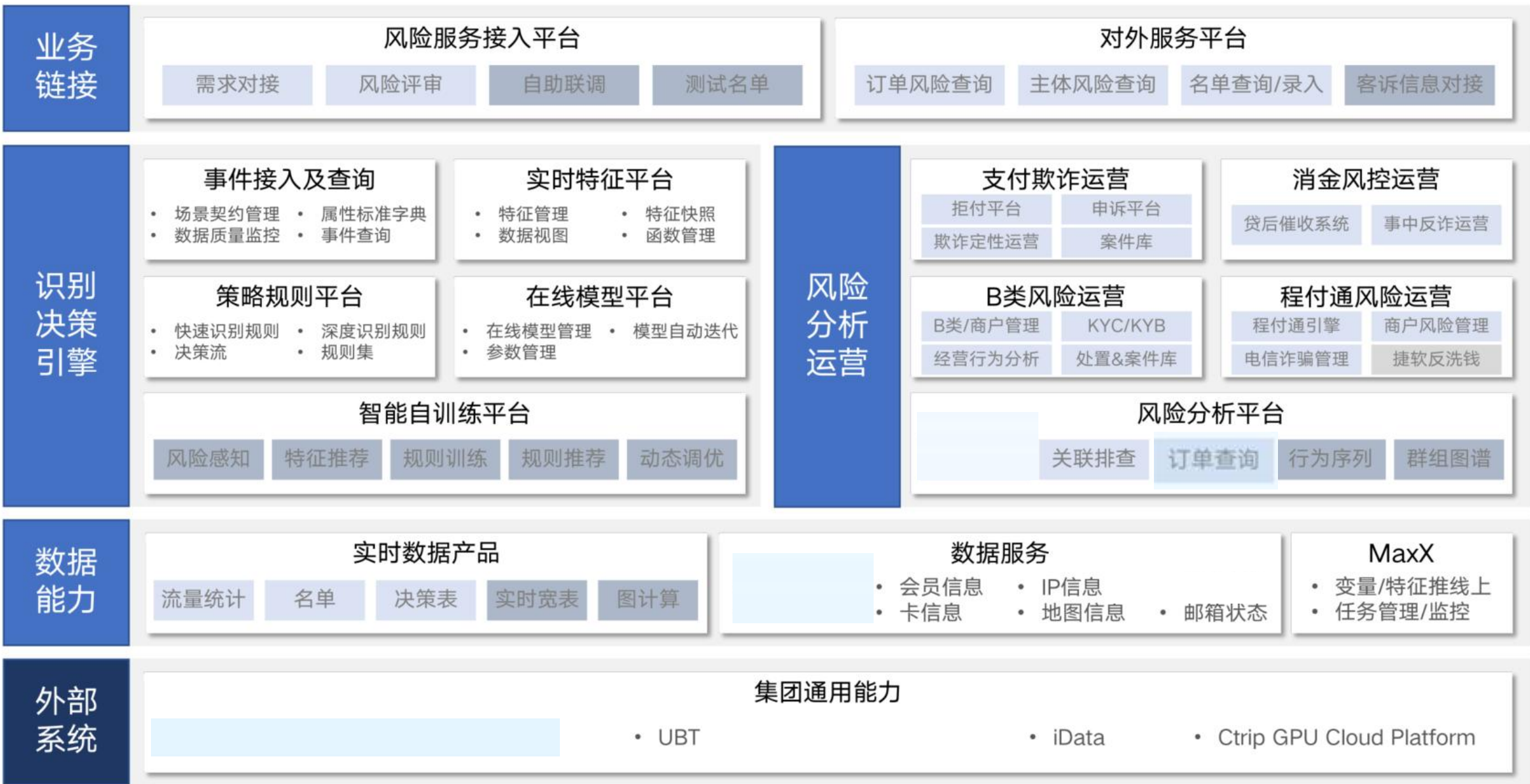
性能卓越
单机QPS和CPU利用率在集团内处于领先水平；

健壮性强
多机房多云DR部署
对突发流量有较快的响应能力，自动扩容

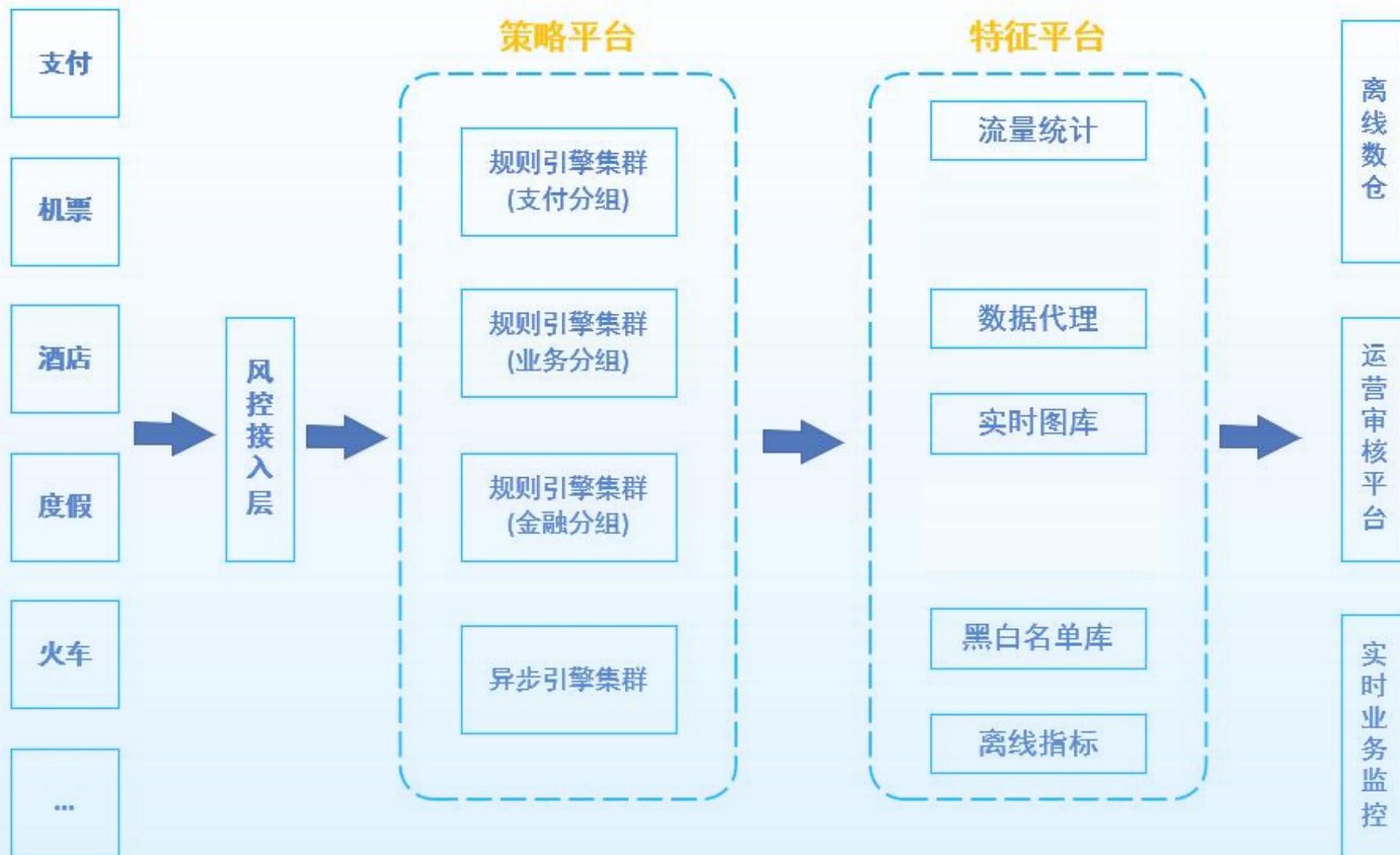
扩展灵活
目前接入了全集团各个BU的风控场景点，活跃的风控场景超过数百个



风控平台组成



风控数据流

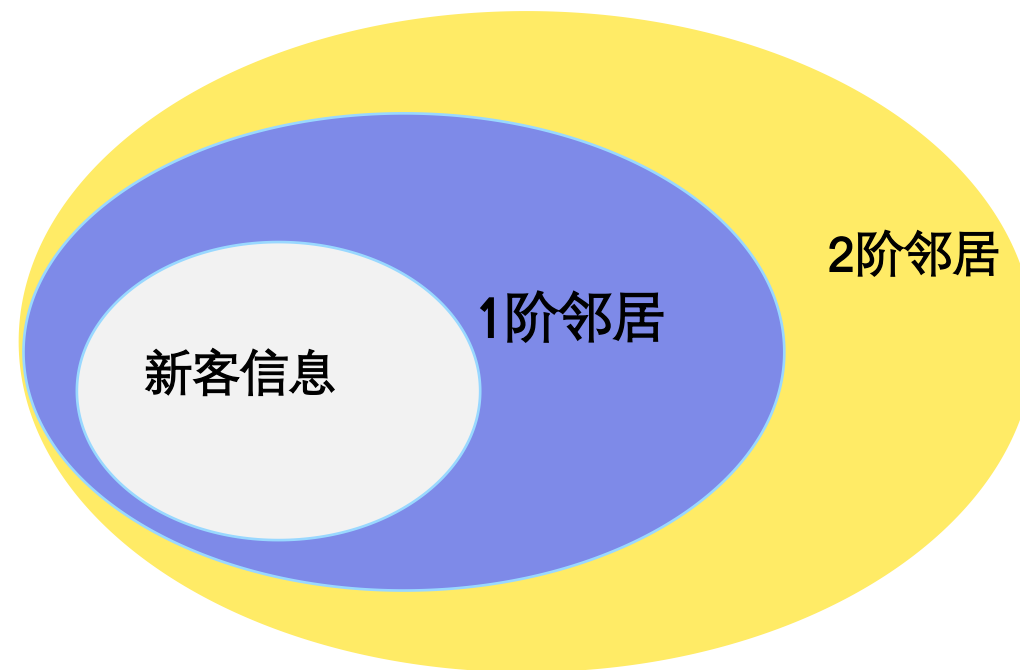


2

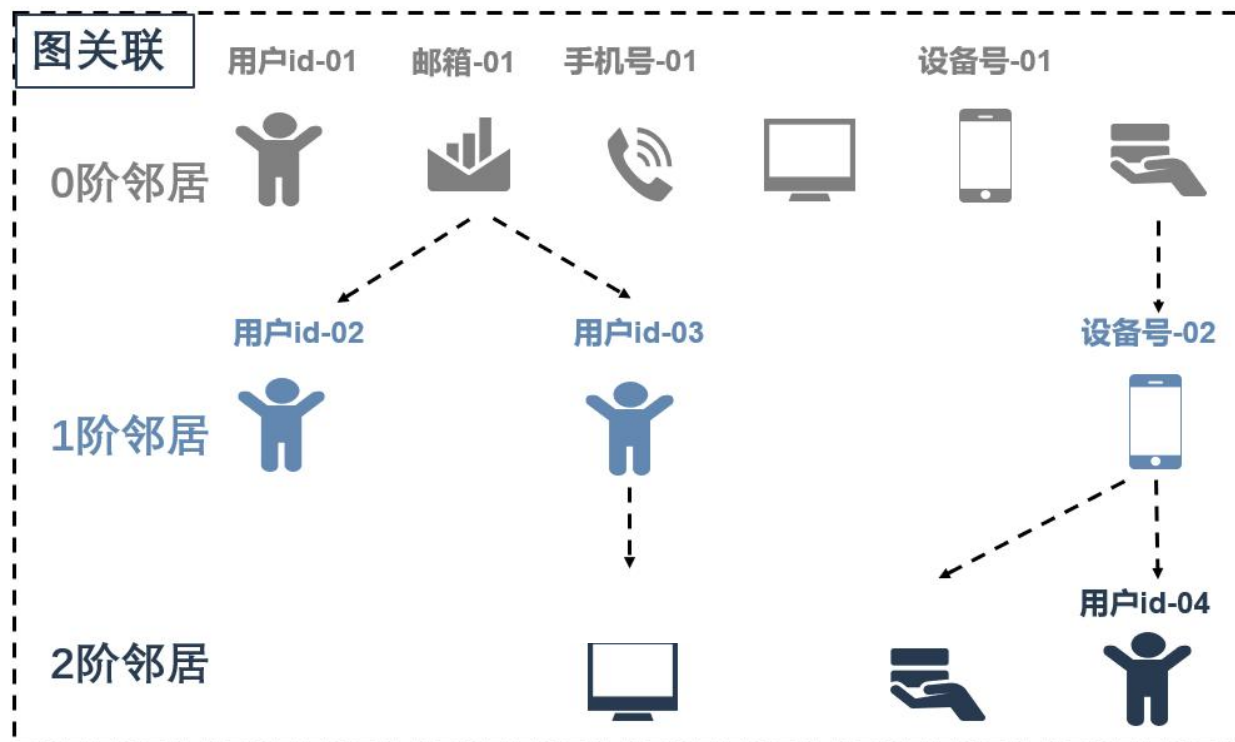
NebulaGraph在携程风控 的落地实践

● 背景

- 携程的海量实时交易中隐藏巨大的欺诈风险
- 新客订单因为历史信息少，
导致欺诈风险是**老客的数倍**
- 用图数据库构建实时动态图，
丰富支付新客的信息，
提高支付订单的**欺诈识别率**



● 业务需求

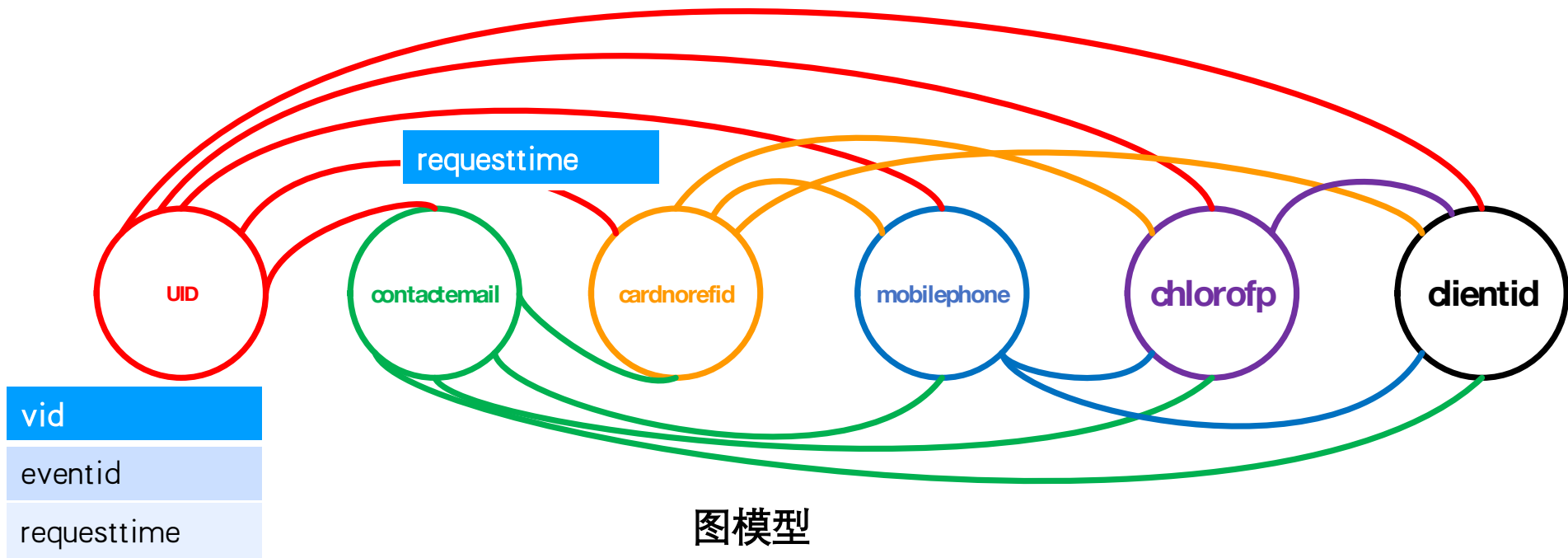


- 对业务行为进行建模, 实时进行图的写入关联
- 支持实时查询当笔交易的**2阶关联**
- 保持**一年内**的实时关联图

图谱建模

原始数据格式

eventid (流水号)	requesttime (订单时间)	uid (用户id)	contactemail (绑定邮箱号)	cardnorefid (卡参考号)	mobilephone (绑定手机号)	chlorofp	clientid (设备号)
b123-xxx-123	2024-07-07 17:56:32	M123xxx456	123xxxmail@q q.com	4352xxxx1234	136xxxx1638	165xxx123	0933xxx1234

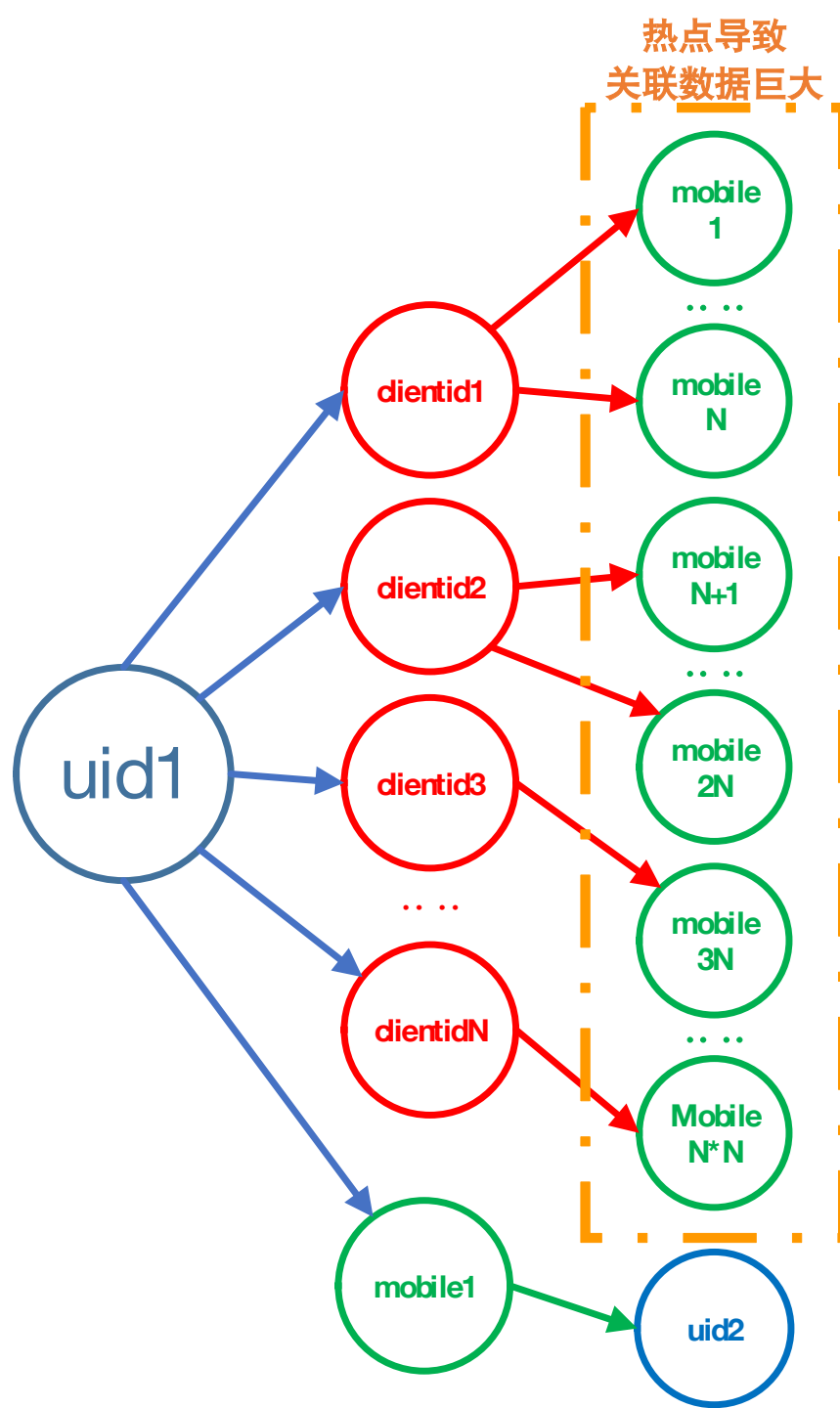


图模型

查询逻辑

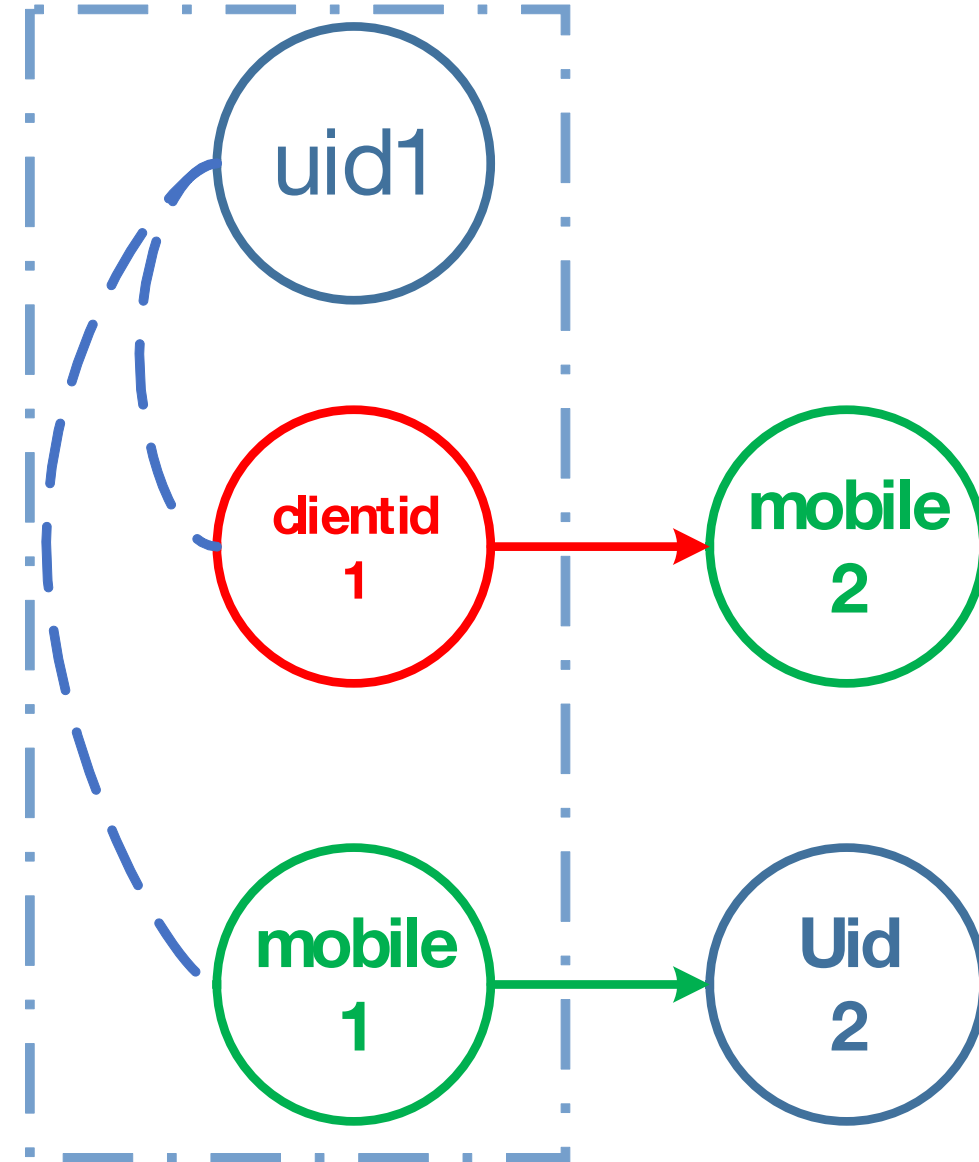
难点1

一阶查询存在**热点**,
导致二阶查询量过大,
发生超时甚至宕机.



● 查询逻辑

当笔交易中的介质元素尚未关联, 无法关联到下阶数据

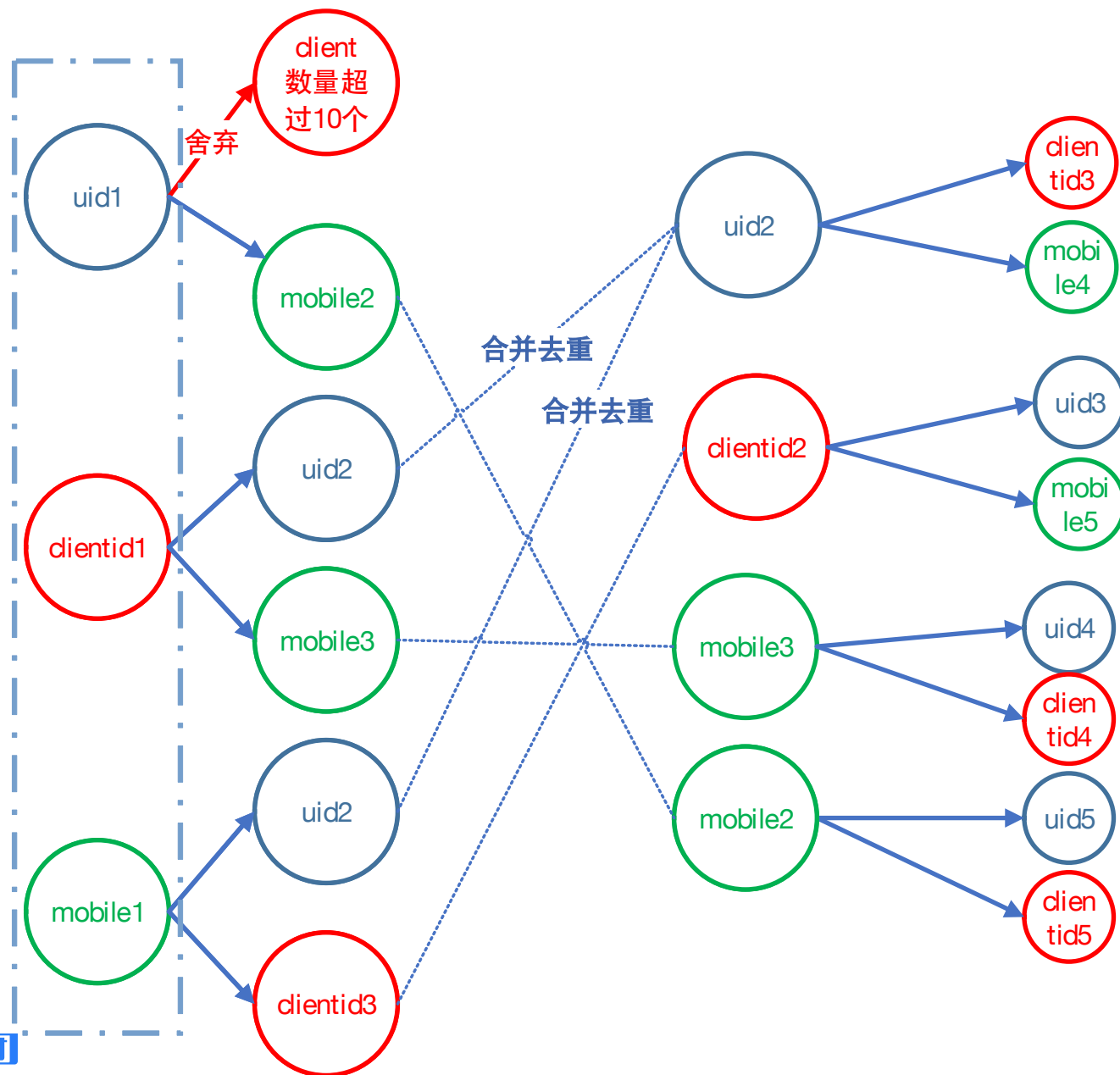


● 难点2

存在查询早于关联写入的情形

查询逻辑

当笔订单元素 一阶关联查询 一阶关联结果整理合并 二阶关联查询 二阶关联结果整理合并



解决

■ **分步查询**, 在第一阶查询时过滤热点.

■ 分类收集一阶关联元素, 去重后查询二阶邻居.

■ 从单个查询改为**群组查询**, 并且使用多线程优化耗时

● 查询逻辑

● 关键NGQL

- 按组查询
- 剔除热点

```
$ MATCH (v1)-[e:link*1]-(v2) where id(v1)in ['WCONK2032748733'] RETURN labels(v2)[0]AS tagName,
CASE size(collect(distinct id(v2))) > 9 WHEN TRUE THEN list[] WHEN FALSE THEN collect(distinct id(v2)) END AS vidSet
```

起始点
['_WCONK2032748733']
热点阈值

Table

tagName	vidSet
clientid	["37001116416966745882", "12001106616247467935", "12001090476626724371"]
mobilephone	[] 剔除热点后返回空
contactemail	["cheelupm123@gmail.com#"]
cardnorefid	["6977430000200239360", "151127023923", "6958816170044358176", "151133094322", "7031507100107016608", "151127020374"]

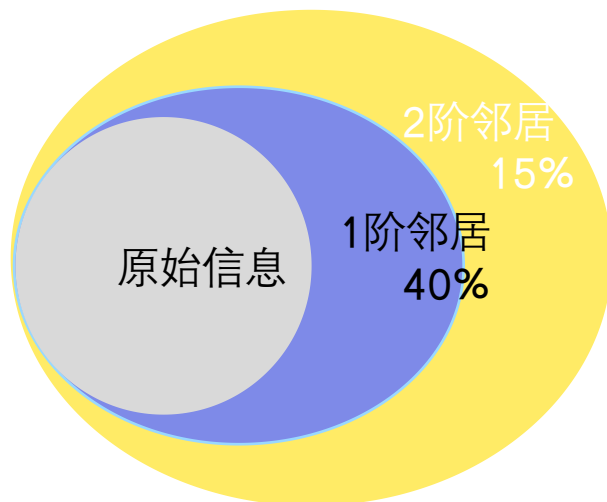
Total 4 <

从起始点开始分类查询关联点, 剔除超过热点阈值的类型点, 即过滤热点

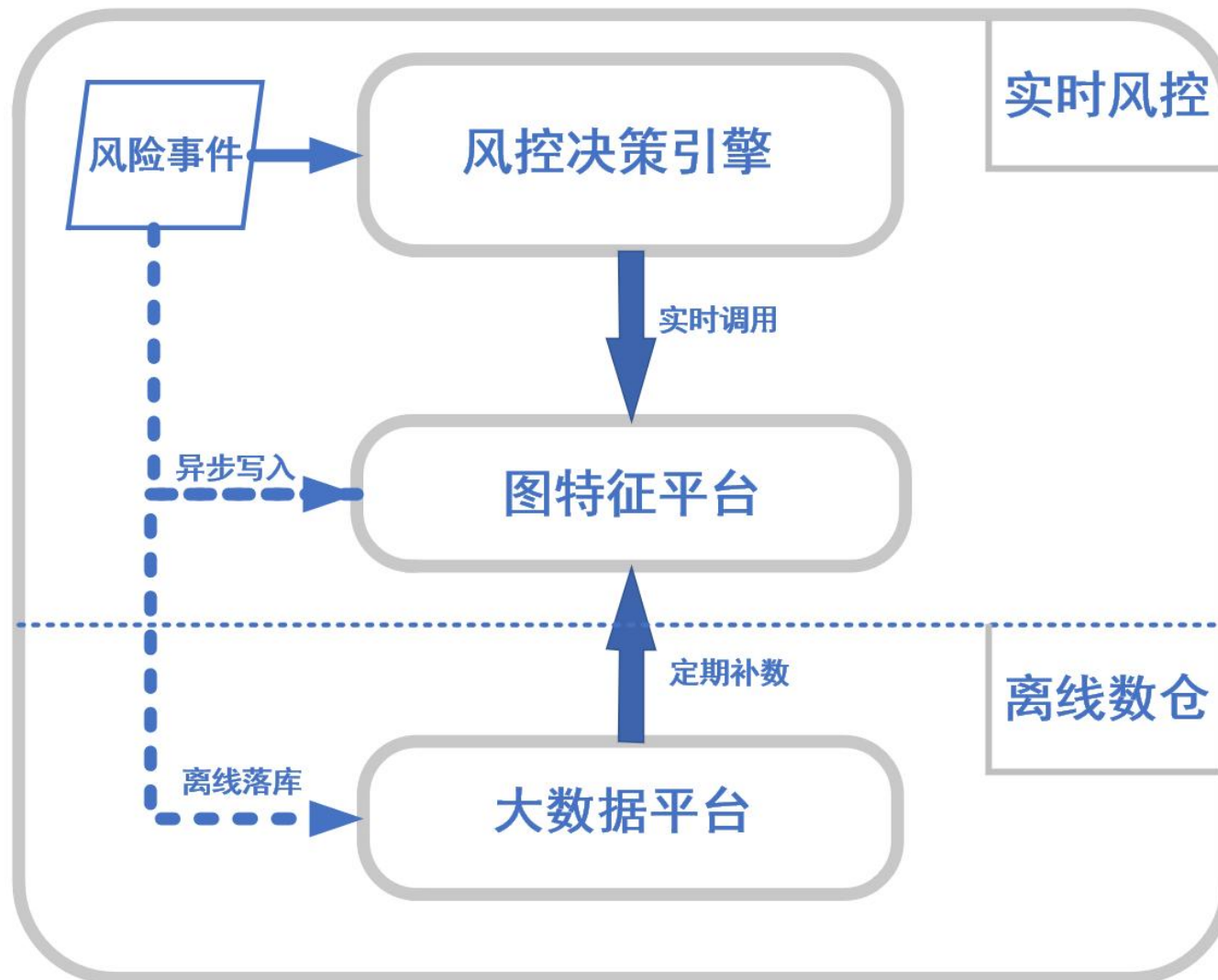
项目成果

关键指标

- 数据量: 千万级的点, 亿级的边
- 耗时: 单次查询平均10ms
- 关联效果: 额外获取55%账户信息
- 业务效果: 欺诈覆盖相对提升32%



实时图特征平台架构



3

总结和展望

● 总结展望

- 图库天然适配风控场景
- 集团主流支持NebulaGraph
- 探索在血缘关系，团伙识别等场景的落地



扫码关注 “携程技术”



在此特别感谢 NebulaGraph 团队和开源社区的技术赋能，愿未来一路同行！

谢谢! Q&A



携程技术公众号

-  <https://nebula-graph.com.cn>
-  GitHub: [vesoft-inc/nebula](https://github.com/vesoft-inc/nebula)
-  Twitter: [@NebulaGraph](https://twitter.com/NebulaGraph)
-  Facebook: [@NebulaGraph](https://www.facebook.com/NebulaGraph)
-  <https://discuss.nebula-graph.com.cn>



微信公众号



开源项目

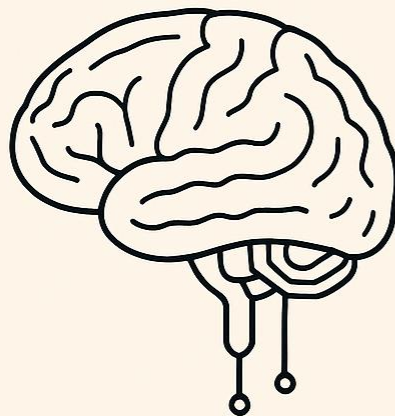
State of NebulaGraph GenAI Team

古思为 Wey GU, NebulaGraph

Domain Knowledge + LLM

Enabling Machines with Domain Knowledge

Fine-Tuning / GaLM



Brain Surgery
/ BMI

RAG / Agentic RAG



Tools for
the Mind

Overview

- Fusion GraphRAG
- Other works recently

Wey GU

> NebulaGraph GenAI Team Lead

> Microsoft MVP(Python, AI)

GenAI, Graph, Open Source, ...

Believe in Open Source & build things with magic
to help people do the same.

-  wey-gu
-  wey_gu
-  weygu
-  siwei.io

RAG

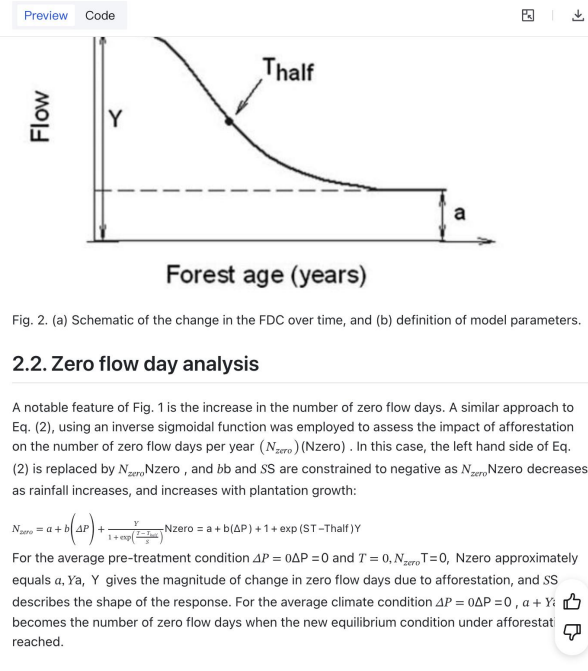
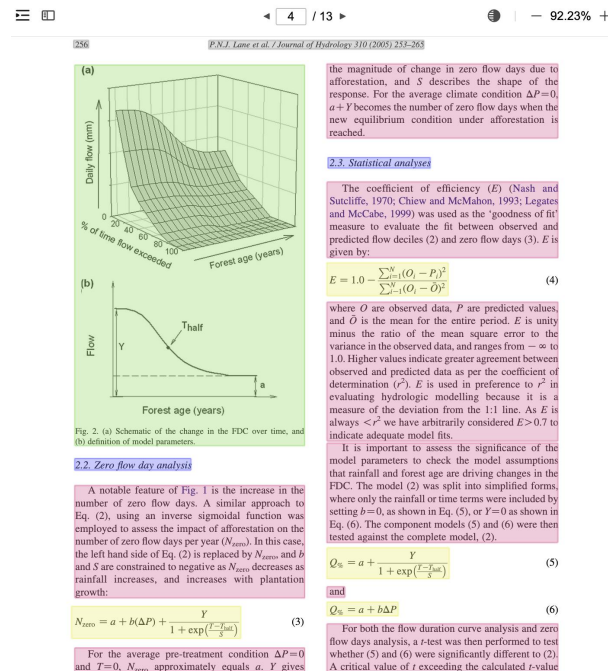


RAG



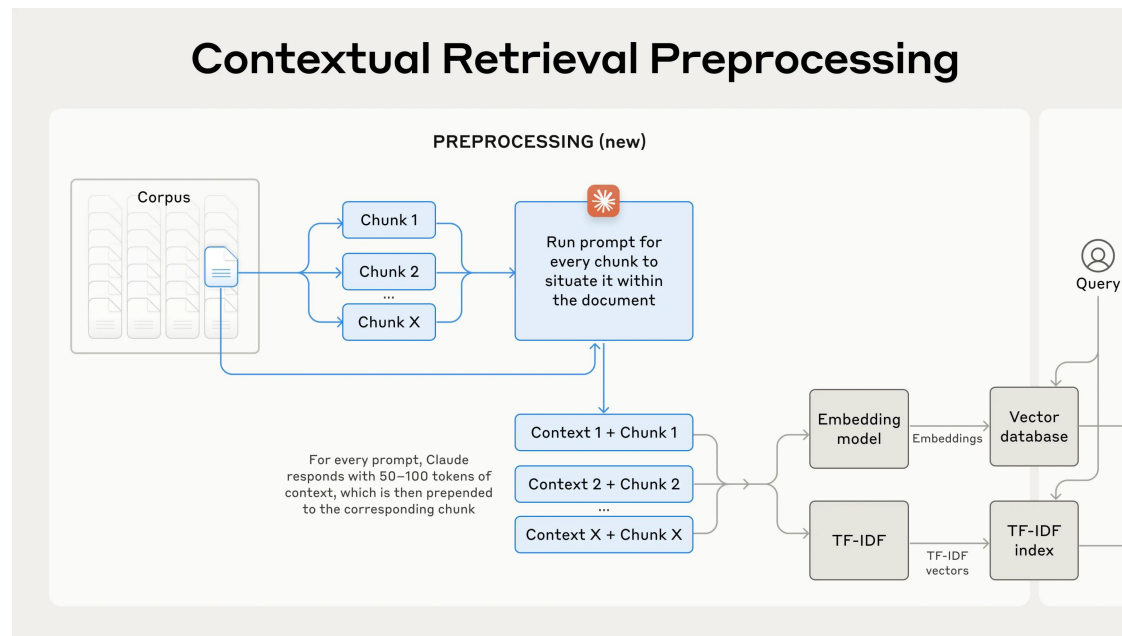
Advanced RAG

- IDP(Intelligent Document Processing)
- Contextual Retrieval
- HyDE(Hypothetical Document Embeddings)
- Rerank
- Tree-RAG
- RAPTOR
- GraphRAG



Advanced RAG

- IDP(Intelligent Document Processing)
- **Contextual Retrieval**
- HyDE(Hypothetical Document Embeddings)
- Rerank
- Tree-RAG
- RAPTOR
- GraphRAG



Advanced RAG

- IDP(Intelligent Document Processing)
- Contextual Retrieval
- **HyDE(Hypothetical Document Embeddings)**
- Rerank
- Tree-RAG
- RAPTOR
- GraphRAG

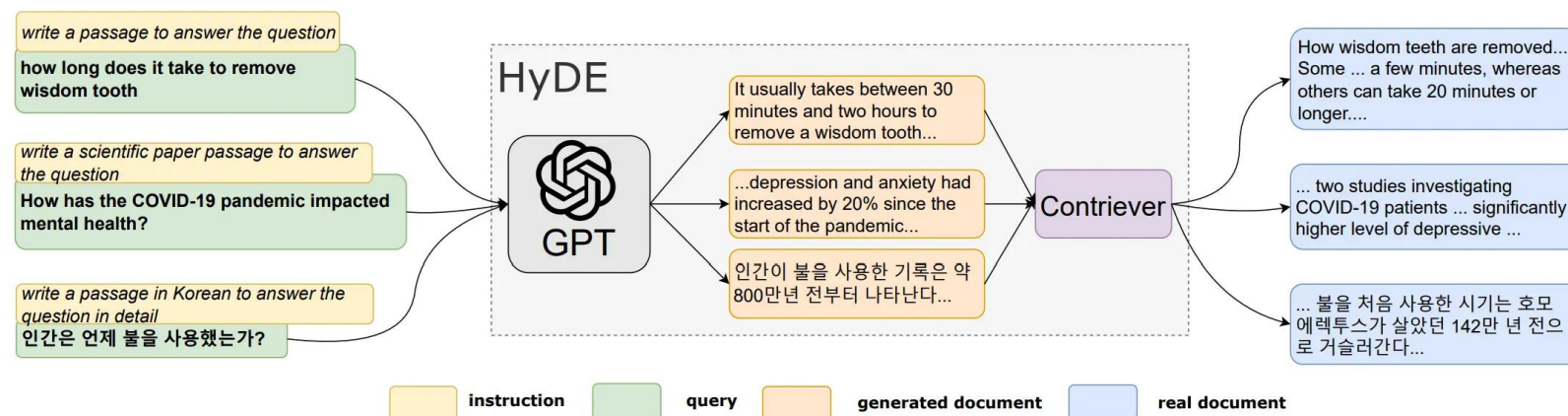
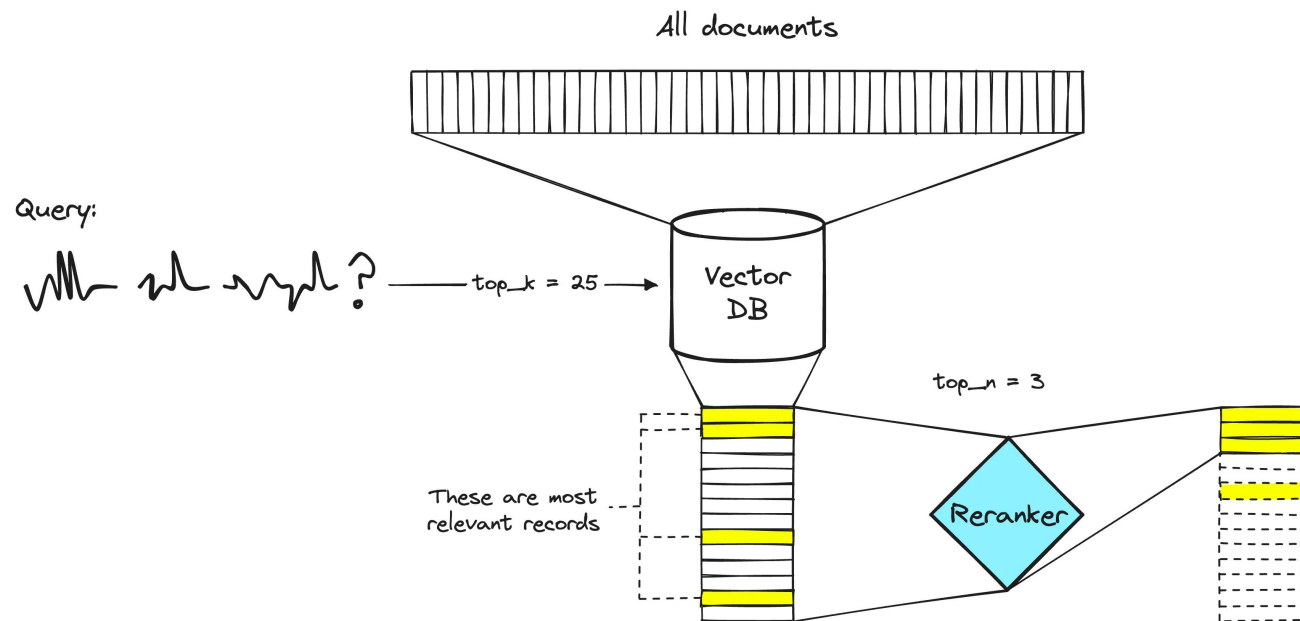


Figure 1: An illustration of the HyDE model. Document snippets are shown. HyDE serves all types of queries without changing the underlying InstructGPT and Contriever/mContriever models.

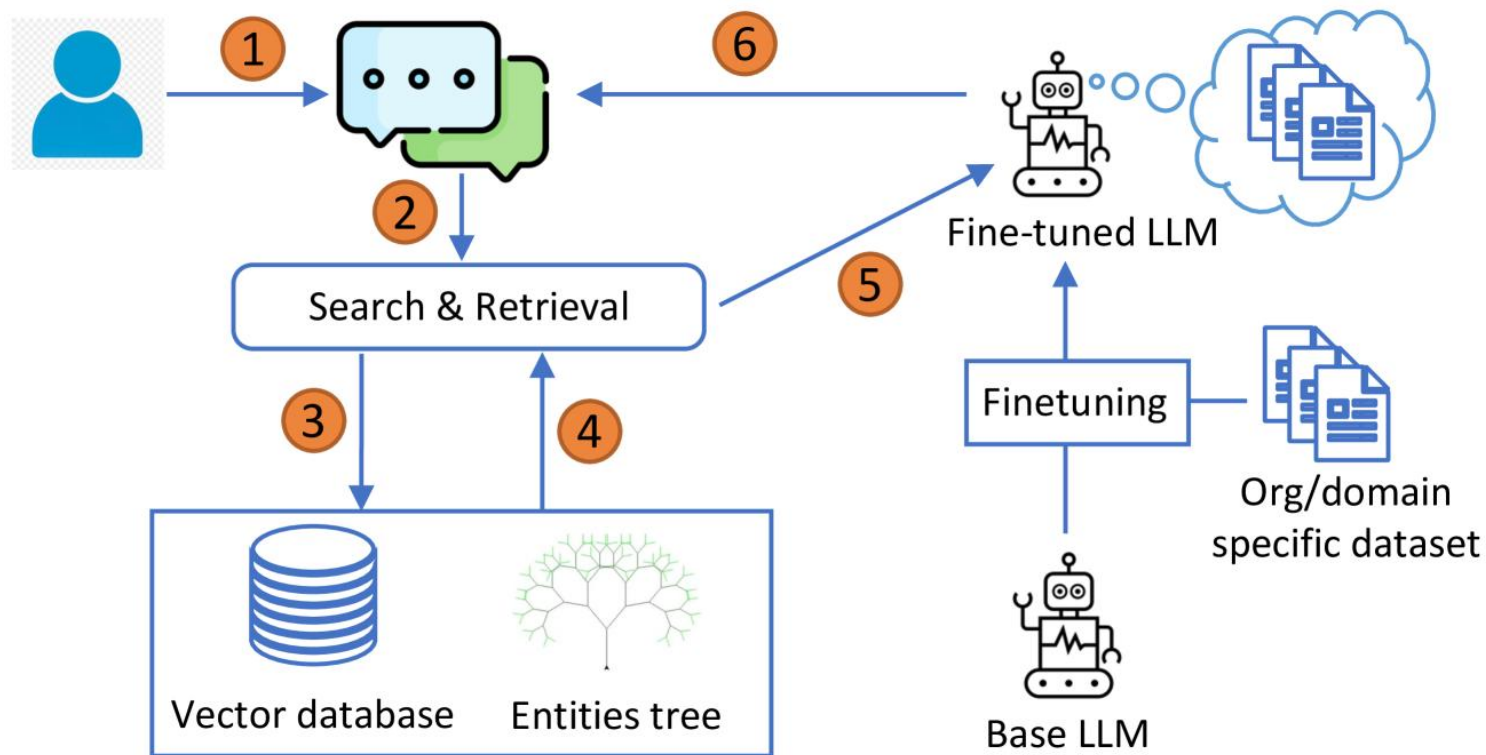
Advanced RAG

- IDP(Intelligent Document Processing)
- Contextual Retrieval
- HyDE(Hypothetical Document Embeddings)
- **Rerank**
- Tree-RAG
- RAPTOR
- GraphRAG



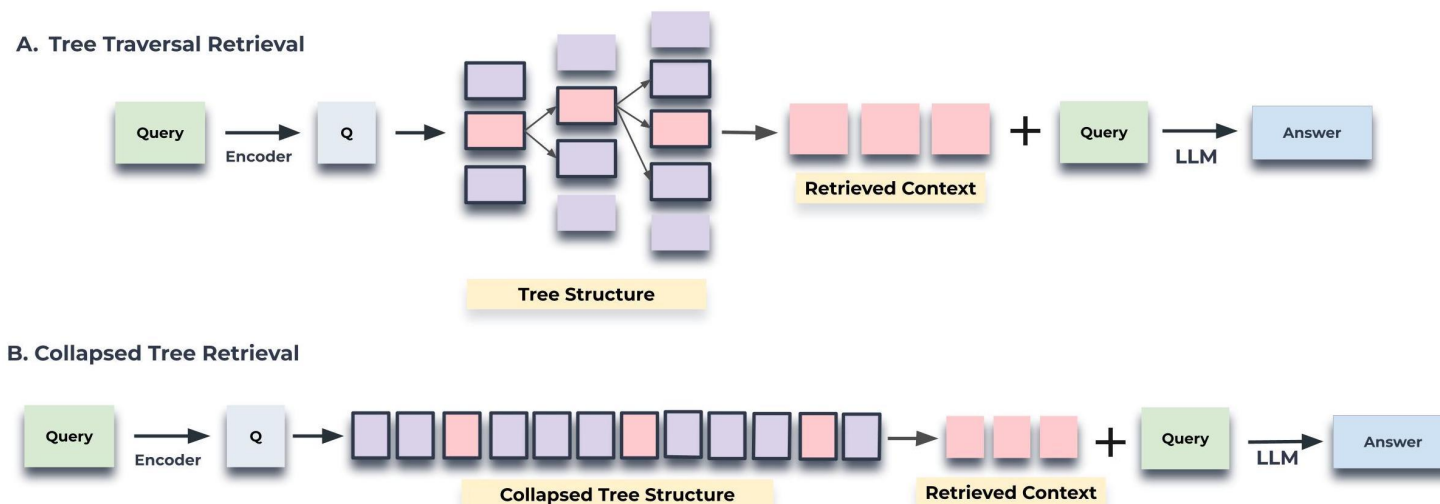
Advanced RAG

- IDP(Intelligent Document Processing)
- Contextual Retrieval
- HyDE(Hypothetical Document Embeddings)
- Rerank
- **Tree-RAG**
- RAPTOR
- GraphRAG

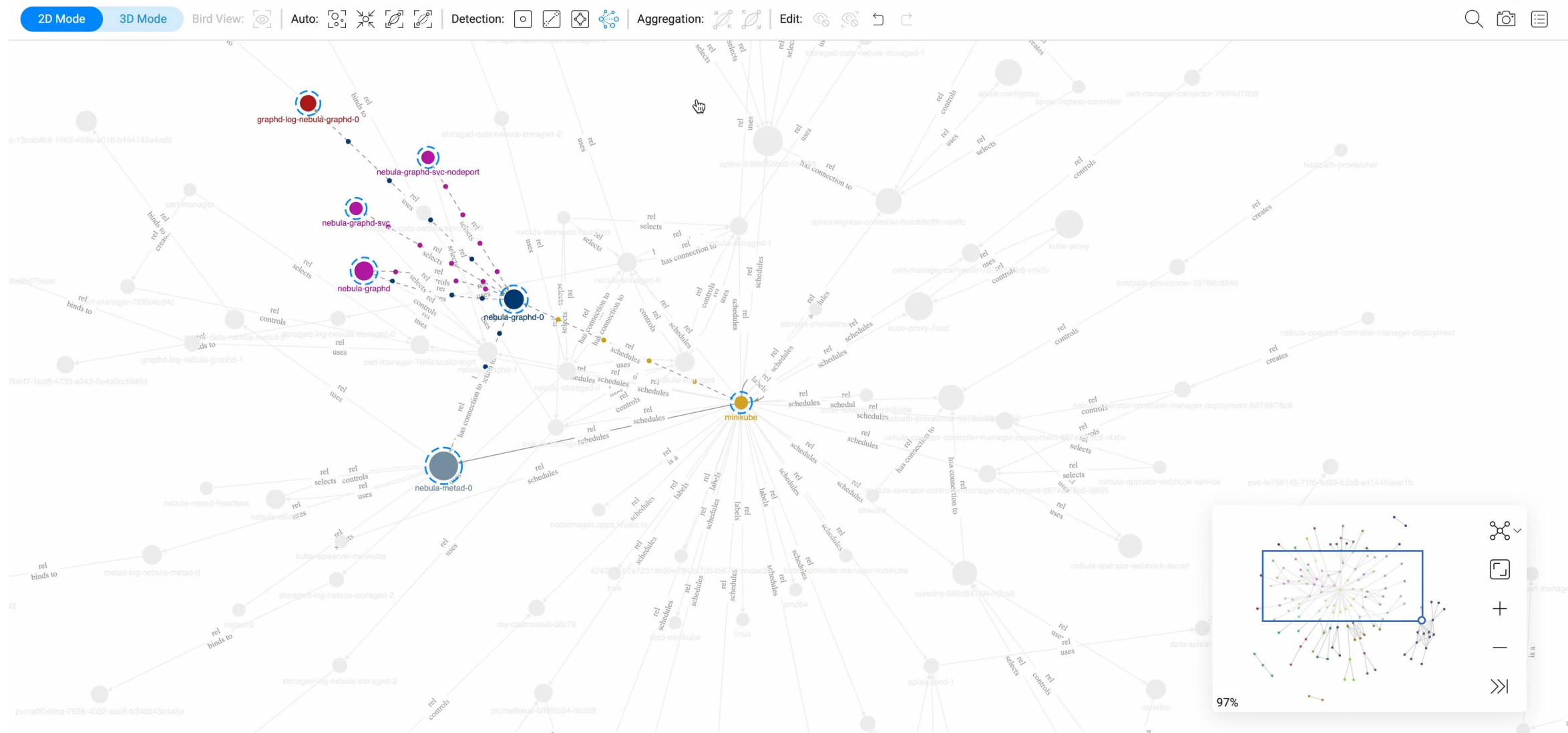


Advanced RAG

- IDP(Intelligent Document Processing)
- Contextual Retrieval
- HyDE(Hypothetical Document Embeddings)
- Rerank
- Tree-RAG
- **RAPTOR**
- GraphRAG



Fusion GraphRAG



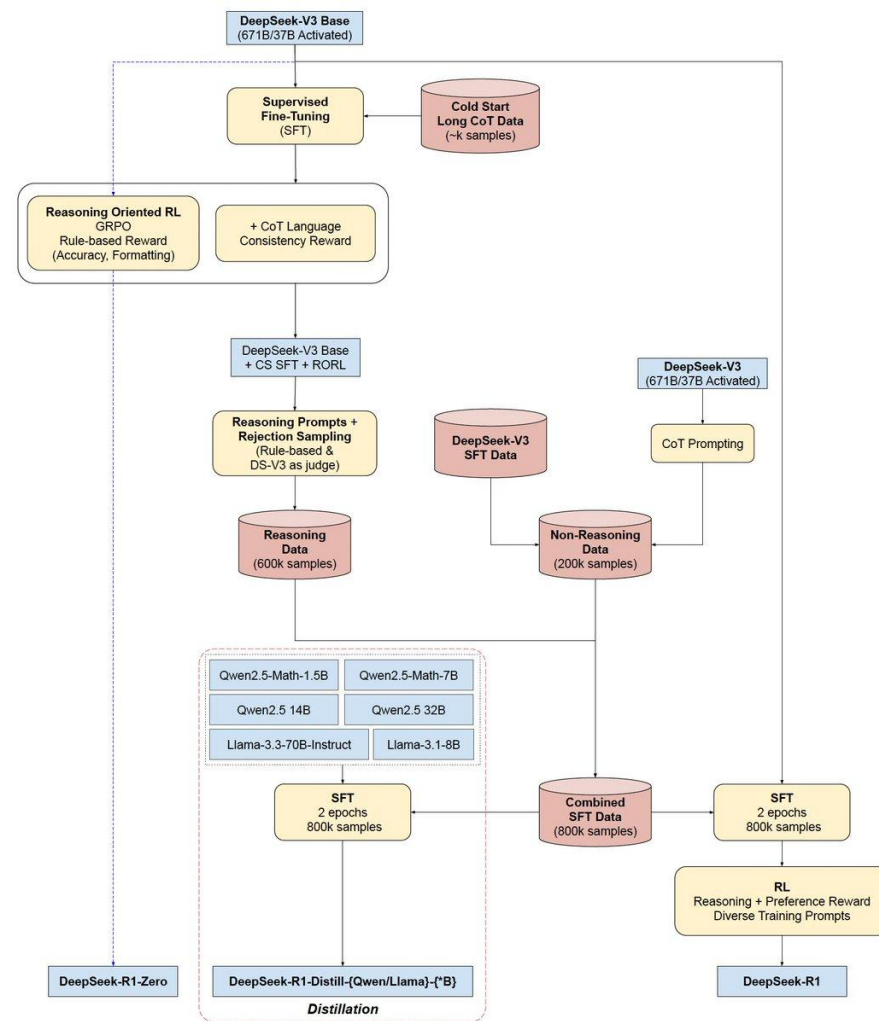
Overview

- Fusion GraphRAG
- Other works recently
 - MCP
 - DeepSeek Inspired Work
- LlamaIndex MCP ToolSpec
 - By PsiACE
- NebulaGraph MCP Server
 - github.com/PsiACE/nebulagraph-mcp-server

Overview

- Fusion GraphRAG
- Other works recently
 - MCP
 - **DeepSeek Inspired Work**
- Reasoning augmented Agentic RAG Planning
- Reasoning augmented Generation
- Reasoning based MetaAgent
- Composer
- Chain of Exploration v2
- GRPO based Graph Extraction

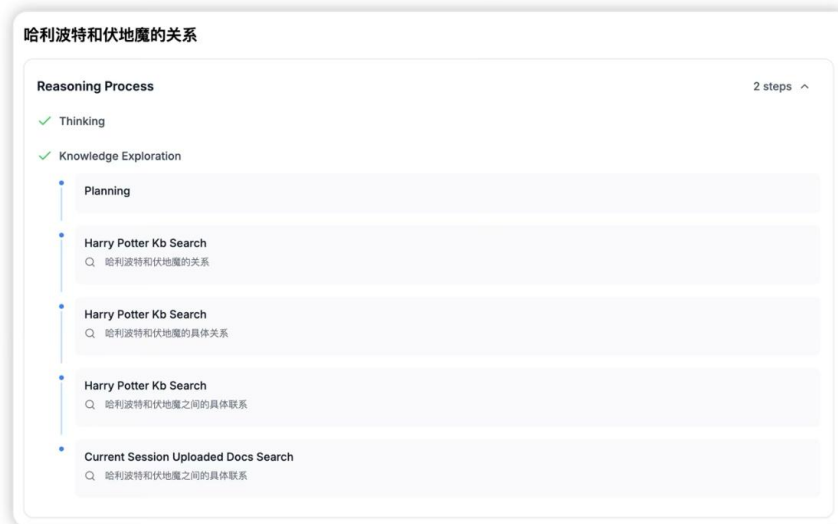
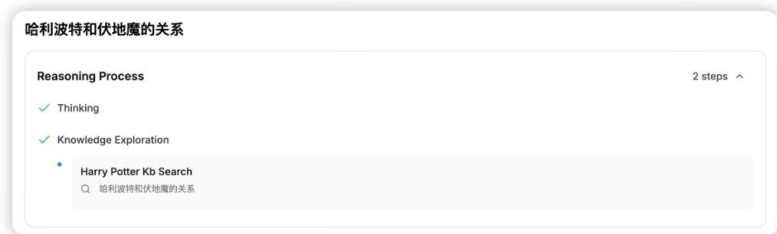
- DeepSeek-coder
- Disk Cache
- DeepSeek-Math/GRPO
- DeepSeek-R1
- DeepSeek-R1-Distill-Qwen2.5-*b
- DeepSeek Janus Pro
- NSA
- 3fs
- V3-0324
- ...



例 1：增强型 Agentic RAG 系统

在 Catalyst 中，我们利用 DeepSeek 的 R1 及 R1-distill 模型作为任务规划器，构建了一个具备高可见性和透明决策过程的 Agentic RAG 系统。

- **改进前：** 传统系统的推理阶段受到用户需求描述和知识桶命名方式的影响，决策过程难以观察。
- **改进后：** 采用 R1 模型后，我们的系统不仅性能大幅提升，而且推理过程完全可视化，用户可以清楚看到模型如何在图、BM25 和向量索引间进行查询和决策。



例 2：生成阶段的深度推理



⊗ Thought for 143 seconds ^

3. 总结与解决办法

- 原因：MATCH 查询是模式匹配，不依赖于边主键中的 ranking；而 FETCH PROP 查询是基于完整的边主键进行查找，默认使用 ranking = 0。当实际插入的边的 ranking 值不是 0 时，用默认的 ranking 查找就会查不到数据。
- 解决方法：确认边在插入时使用的 ranking 值，如果不是 0，则在 FETCH PROP 查询中需要加上正确的 ranking。比如，如果边的 ranking 为 5，则查询语句应写为：

```
ngql
FETCH PROP ON edgeA "s1" -> "t1"@5 YIELD properties(edge) AS e;
```

这样，FETCH PROP 就能精确匹配到那条边了。

另外，还要注意：

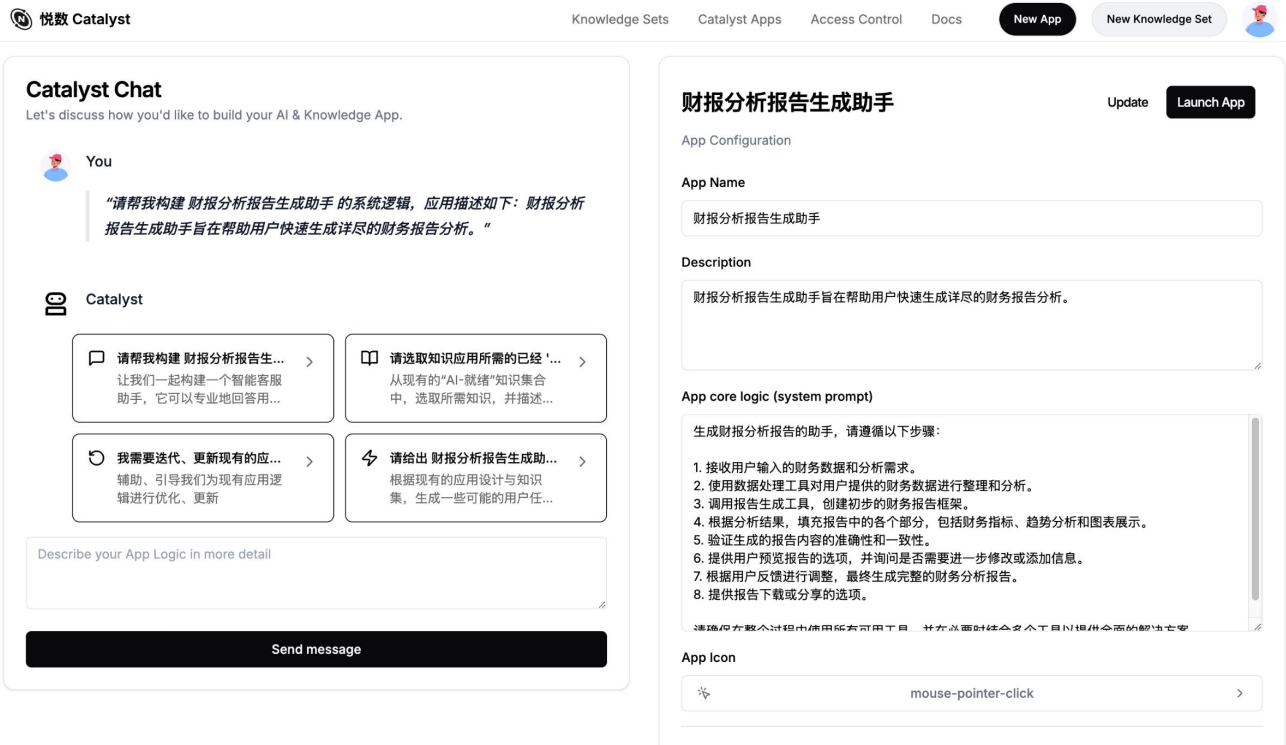
- 在 NebulaGraph 中，字符串常量应当使用单引号（如 's1'）书写。虽然某些情况下双引号也能使用，但为了避免解析上的歧义，推荐使用单引号。
- 请确认 schema 定义和数据写入时是否有设置非默认的 ranking 值。

综上所述，MATCH 能返回结果而 FETCH PROP 返回空的主要原因就在于边的 ranking 值没有在 FETCH PROP 查询中指定，导致默认使用的 ranking 值 0 与实际数据不匹配。

例 3:MetaAgent 辅助应用生成

在部分场景下，用户提供的信息极少，导致初始应用逻辑混乱甚至无效。

- 改进措施： 利用 R1 模型在推理阶段扩展思考深度，自动生成并优化初始应用逻辑，确保输出结果具有较高的合理性和实用性。

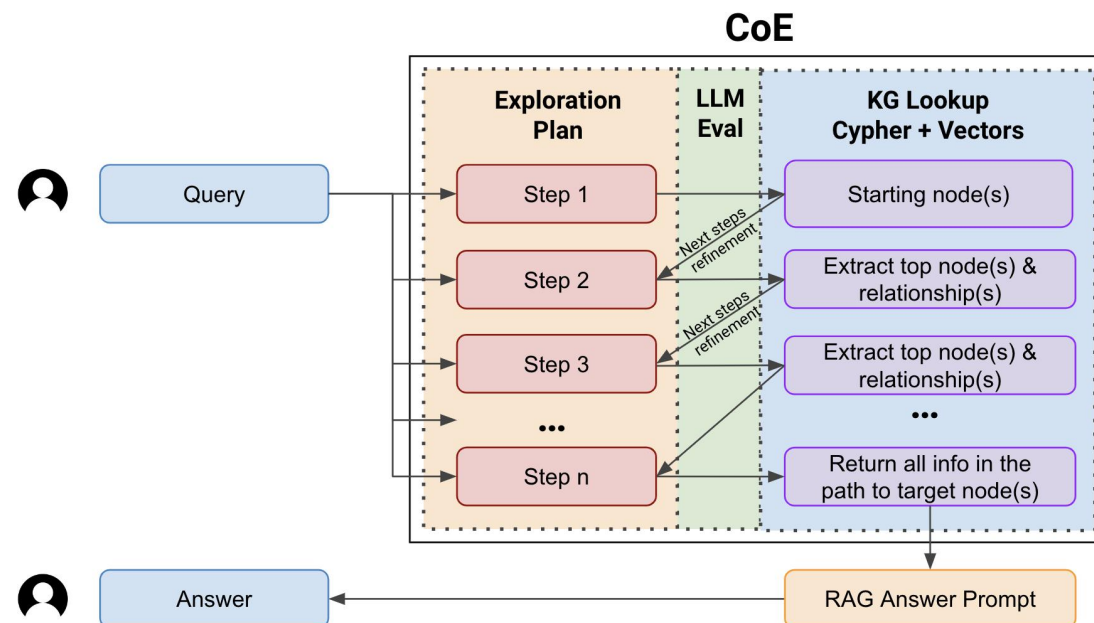
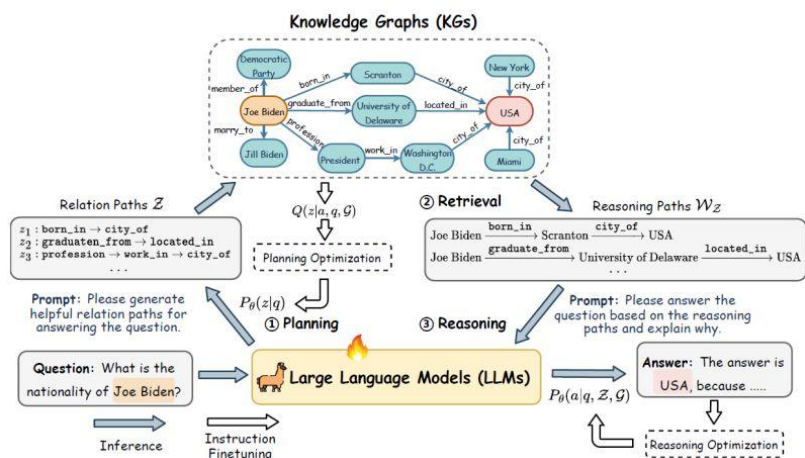
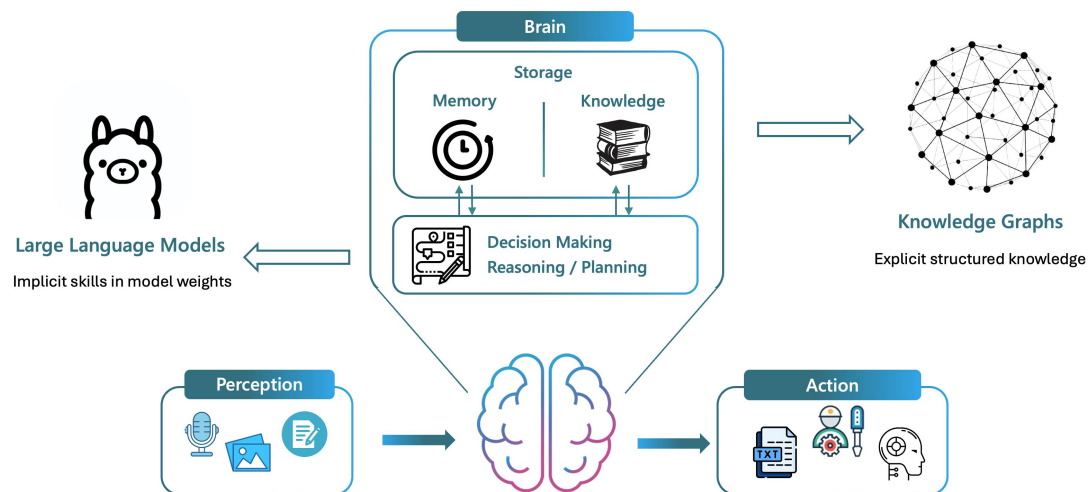


例 4：报告/文章生成器

Catalyst 中的一类应用为报告与文章生成器。

- 应用特点：** 通过可选的推理规划模块，R1 模型在生成初稿时大幅提升内容质量，并在后续迭代中不断优化逻辑链和细节，显著提升生成报告的完整性和专业性。
- 深度研究：** 此外，我们的深度研究功能借助 R1 进行多轮规划和研究，生成高质量、经过深度验证的输出。

例 5: Chain of Exploration

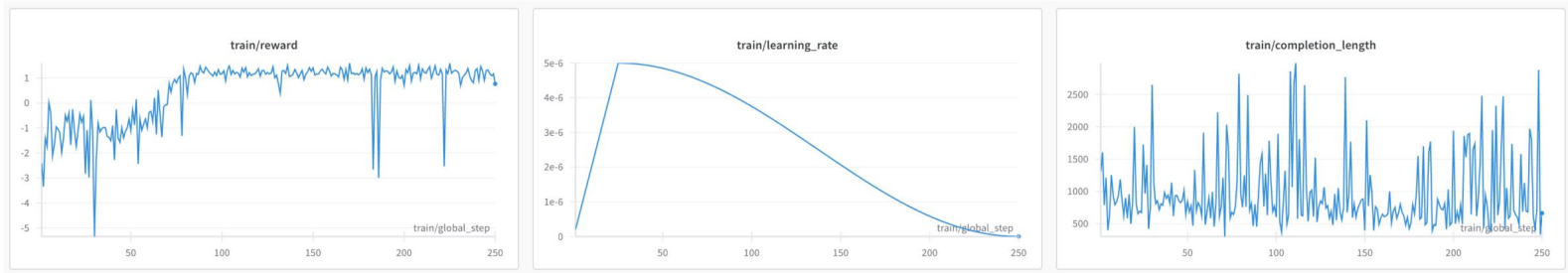


Refs:

- <https://www.linkedin.com/feed/update/urn:li:activity:7198589556848263168/>
- <https://arxiv.org/abs/2405.12035>
- <https://seucoin.github.io/workshop/llmkg/>

例 6：知识蒸馏、知识抽取

Criterion	qwen2.5-3b	qwen2.5-3b-GPRO	gpt-4o-mini
Accuracy	★★★★☆	★★★★☆	★★★★☆
Completeness	★★★★☆	★★★★★	★★★★☆
Coherence	★★★★☆	★★★★☆	★★★★☆
Relevance	★★★★☆	★★★★☆	★★★★☆



github.com/wey-gu/grpo-graph-extraction

Recap

- Fusion GraphRAG
- Other works recently
 - MCP
 - DeepSeek Inspired Work

Wey GU

> NebulaGraph GenAI Team Lead
> Microsoft MVP(Python, AI)

GenAI, Graph, Open Source, ...

Believe in Open Source & build things with magic

to help people do the same.

-  [wey-gu](#)
-  [wey_gu](#)
-  [weygu](#)
-  [siwei.io](#)

- LlamaIndex MCP ToolSpec
 - By PsiACE
- NebulaGraph MCP Server github.com/PsiACE/nebulagraph-mcp-server
- Reasoning augmented Agentic RAG Planning
- Reasoning augmented Generation
- Reasoning based MetaAgent
- Composer
- Chain of Exploration v2
- GRPO based Graph Extraction github.com/wey-gu/grpo-graph-extraction

谢谢! Q&A



<https://nebula-graph.com.cn>



GitHub: [vesoft-inc/nebula](https://github.com/vesoft-inc/nebula)



Twitter: [@NebulaGraph](https://twitter.com/NebulaGraph)



Facebook: [@NebulaGraph](https://www.facebook.com/NebulaGraph)



<https://discuss.nebula-graph.com.cn>



微信公众号



开源项目