

# GraphRAG + 教育



NebulaGraph | Community



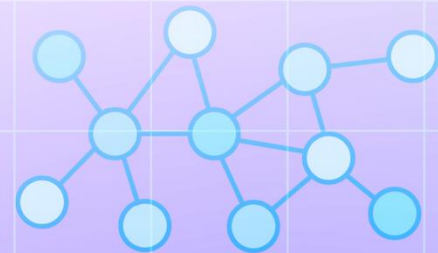
南京大学信息管理学院

# 产学研合作官宣

# 图谱新篇

2025 年 1 月 11 日

南京大学仙林校区信息管理学院





# 信息管理學科

## 與圖數據、知識管理的不解之緣

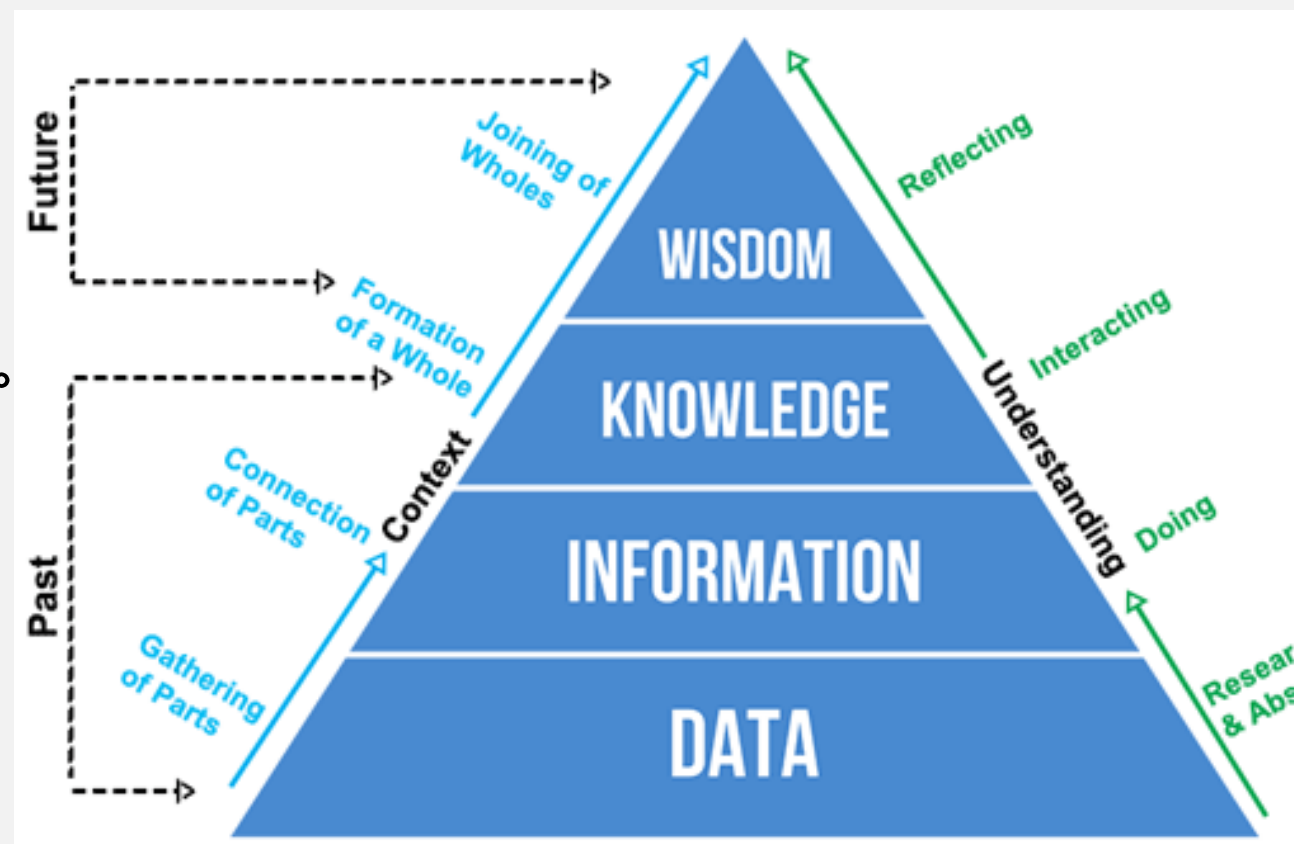
顏嘉麒

2025. 1. 11

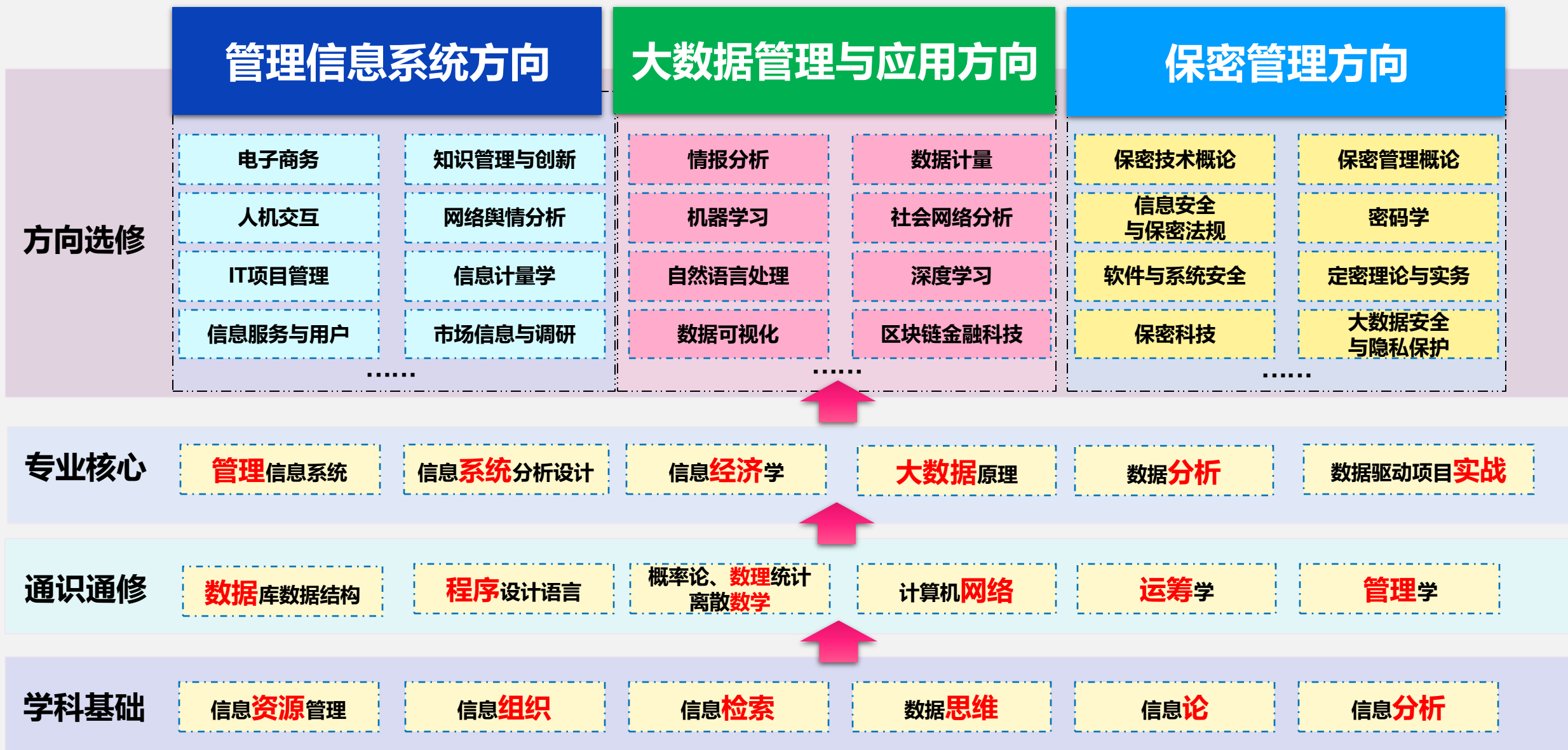


## 信息管理学科与图数据/知识管理的联系

- 信息管理是一个综合性领域，它专注于信息的有效收集、组织、存储、检索和利用。
- 知识管理是信息管理学科中的重要一环。
- 图数据，作为一种特殊的数据结构，由节点和边组成，用于直观地表示实体及其之间的关系。在信息管理中，图数据的应用极大地丰富了信息的组织形式和呈现方式。
- 图数据为信息/知识管理提供了直观、高效的信息组织结构。
- 信息管理学科有很多课程是与图数据/知识管理相关的。



# 信息管理与信息系统专业课程体系



# 南大的信息管理与信息系统

## 南大信管一流的学科与一流专业

### 学科发展：

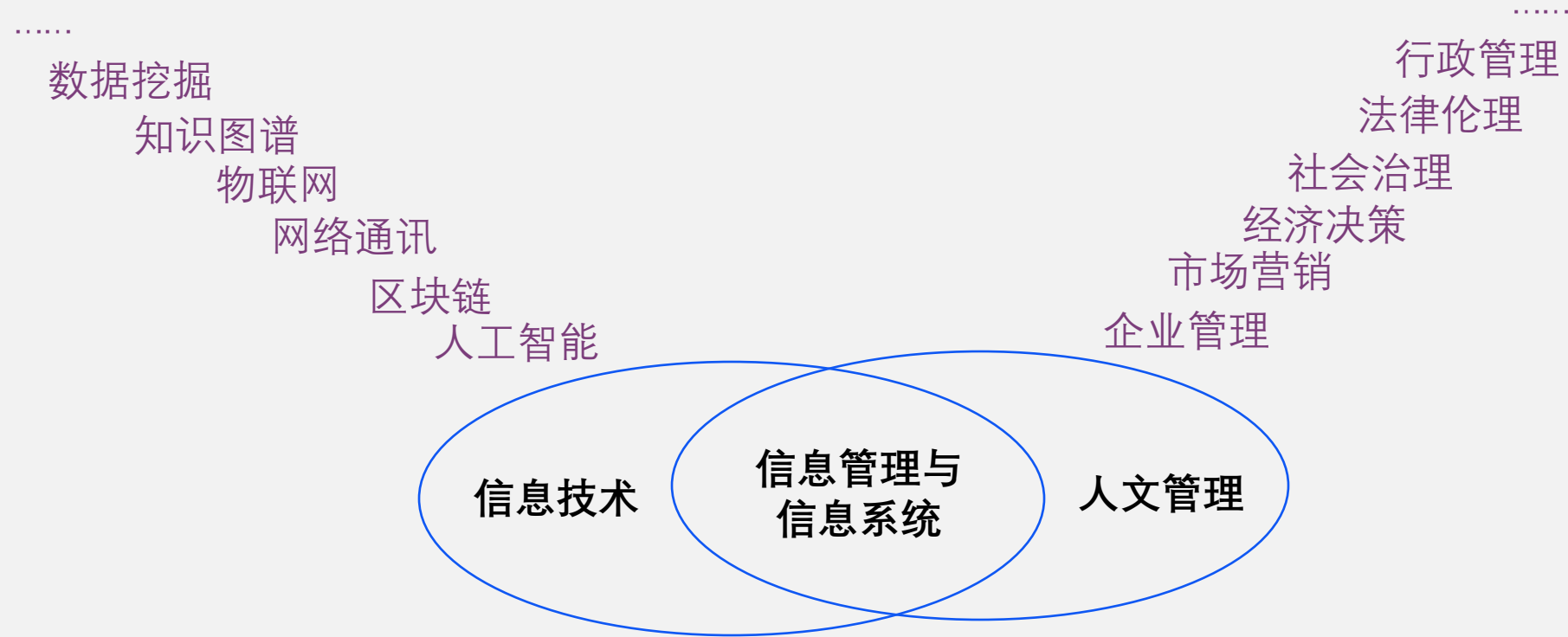
- 2021、2017年，所在学科“信息资源管理”连续入选**国家“一流”**学科建设名单
- 2022、2017年，连续两次在**教育部**学科评估中，获得**最优异**的评级

### 专业建设：

- 2019年，信息管理与信息系统专业被评为**一流本科专业**培育序列
- 2020年，信息管理与信息系统专业入选首批**国家一流**本科专业



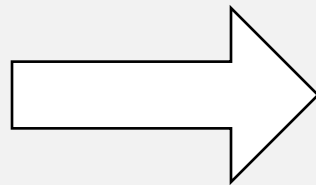
# 数智时代的信息管理与信息系统



- 关心**信息**技术的发展，同样关心技术发展背后的**管理**。
- 信息领域的发展，永远不仅仅是**技术**问题，更是与“**人**”的管理问题紧密相连。

# 一流科研培养一流人才

一流科研水平



一流人才培养

## 南京大学信息管理学院研究中心/团队

- 教育部“数据智能与交叉创新”重点实验室（培育）
- 江苏省数据工程与知识服务重点实验室
- 国家级互联网+双创示范基地
- 南京大学——腾讯互联网+研究中心
- 南京大学人文社会科学大数据研究院
- 南京大学国家保密学院
- 政务数据资源研究所
- 国家信息资源管理南京研究基地
- 南京大学信息技术开发研究所

# 一流科研培养一流人才

## ◆一流科研平台，引领一流本科教育



### 教育部哲学社会科学重点实验室

“南京大学数据智能与交叉创新实验室”（培育）



### 中宣部国家新闻出版署重点实验室

“智慧出版与知识服务重点实验室”



### 省部级重点实验室

江苏省数据工程与知识服务重点实验室  
中国纪念馆协同研究中心  
国家档案局重点实验室（申报中……）



### 江苏省高校国际合作联合实验室

“信息学国际合作联合实验室”



# 国际化视野培养一流人才

## 国际化平台



南京大学-剑桥大学  
国际联合信息与社会研究计划



INDIANA UNIVERSITY  
BLOOMINGTON  
南京大学-印第安纳大学



南京大学-南澳大学



南京大学-匹兹堡大学



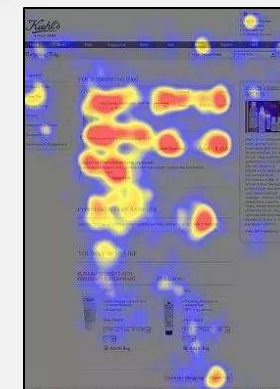
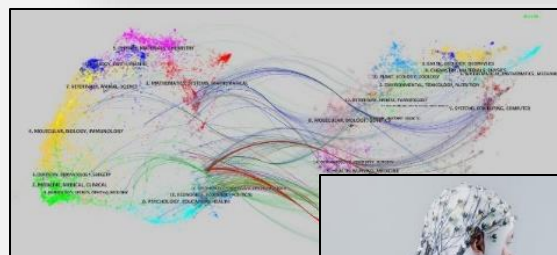
ILLINOIS  
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN  
南京大学-伊利诺伊大学  
信息科学联合实验室



南京大学-德雷塞尔大学



University of  
South Australia  
南京大学-谢菲尔德大学



**信息学国际合作联合实验室**是江苏省“十四五”首批高校国际合作联合实验室，是学科国际合作与跨学科研究的突破性成果。联合实验室通过与伊利诺伊大学，主要开展**定量信息参数测量**、**心脑信息联合分析**、**信息行为实验研究**等。

# 毕业去向

## 升造

## 就业



**欢迎加入南京大学信息管理与信息系统专业**

# NebulaGraph Community

NebulaGraph PMC, Sherman

2025.01.11

# 目录

**1** 图和图数据库

**2** NebulaGraph Community

**3** 高校合作计划，校园大使



# 1 图和图数据库

## 图数据库市场规模



2018



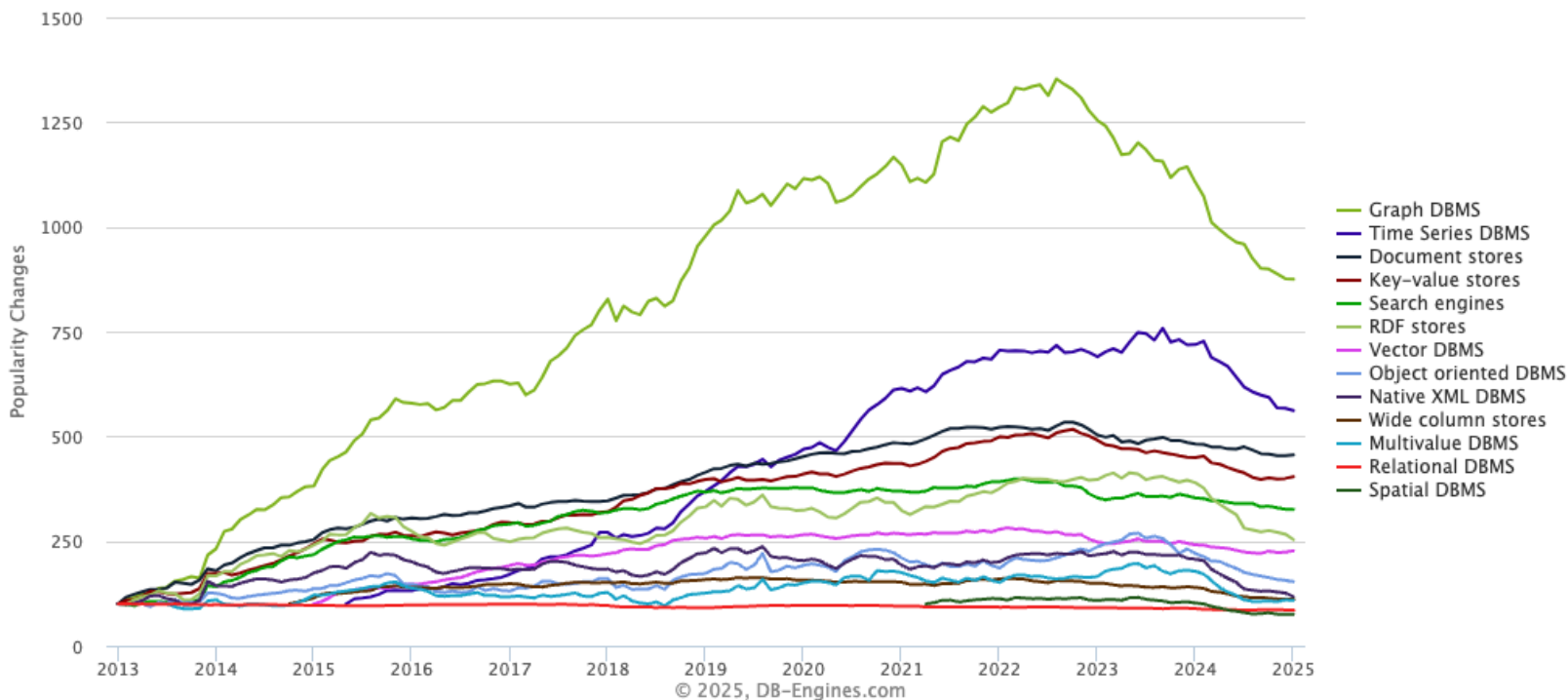
2024-2026



2026

# 图数据库趋势

Complete trend, starting with January 2013

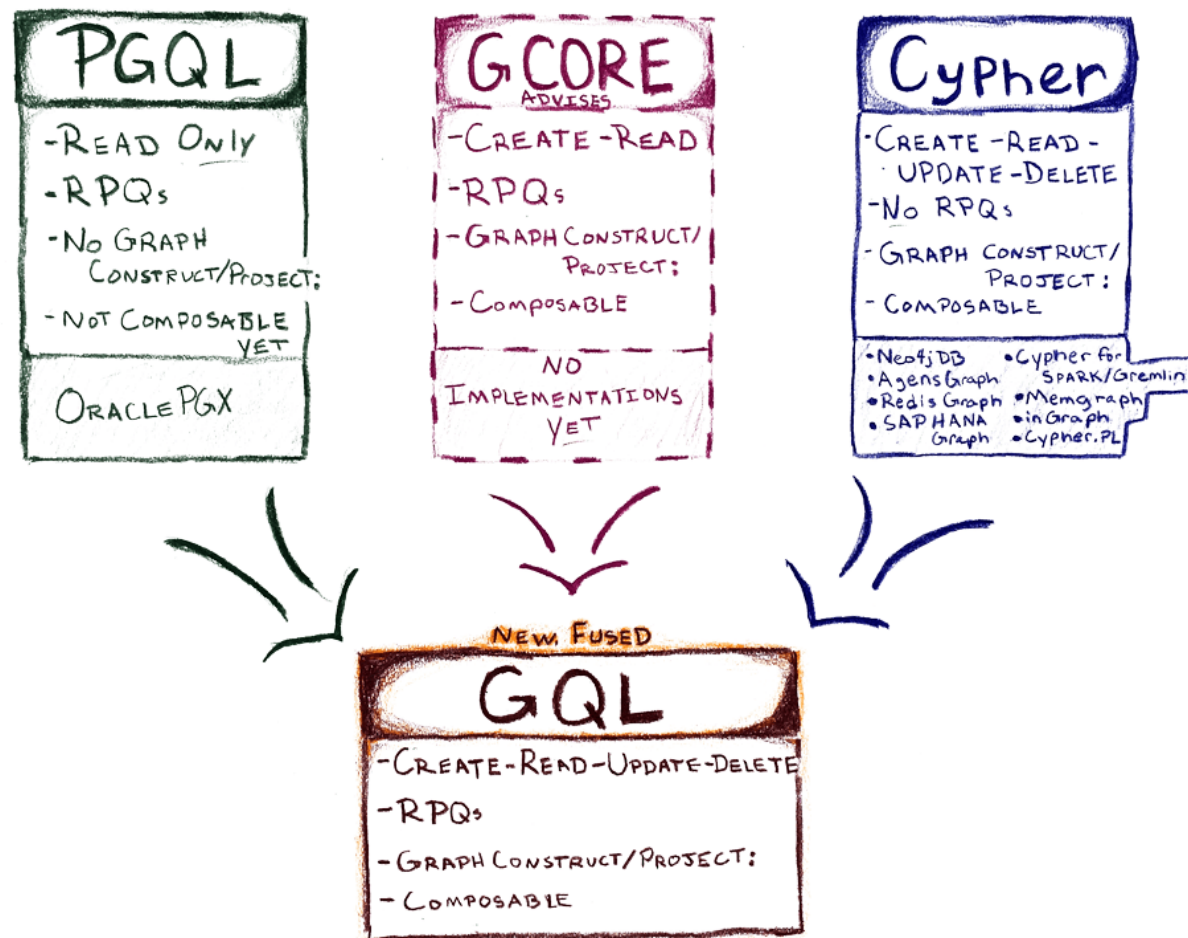


DB-Engines 是全球知名的数据库流行度排行榜网站

# ISO-GQL

## 全面支持 ISO-GQL v1.0

- ❑ GQL 2024.4月 发布，第二门国际数据库标准语言
- ❑ GQL 是完整的数据库语言，包含了 DDL、DML、DQL等
- ❑ GQL 又不仅仅是一门数据库语言，还明确了图数据库相关概念的完整定义
- ❑ GQL 相比当前的图数据库查询语言的关键特性：
  - GQL-catalog & GQL-data分离
  - 支持强 Schema 类型
  - 从简单模式匹配到复杂模式匹配
  - 从单图到多图
  - 支持无向图




# 应用场景

反欺诈  
反洗钱  
信用积分

担保圈识别  
企业图谱


金融证券



实时推荐  
数据治理  
智能问答

服务依赖分析  
代码依赖分析  
智能运维  
智能助理


互联网



知识图谱

疫情·密接分析  
反恐刑侦


政府



生产流程管理  
质量追溯

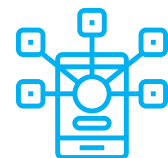
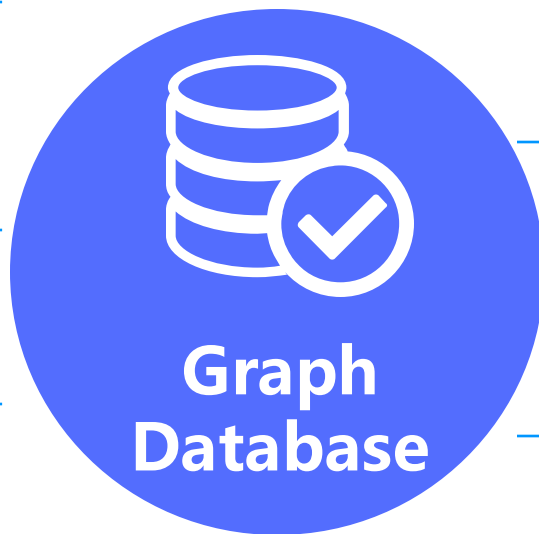
供应链管理  
风险评估  
组织架构协作

制造



电力 燃气 智慧家庭

物联网

## 商业智能

基于图的商业智能解决方案从数据关系的角度去做探索，找到事物之间隐藏的联系，挖掘数据价值



## 人工智能

通过图分析补充数据间关系的特征，提高人工智能做出决策的准确性和效率



## 机器学习

图可以通过边连接和遍历数据，为机器学习实时提取图特征，提高结果准确性



# Graph + AI

从原始数据到使用图技术成本较高：

- **建图成本高**：通过抽取完成非结构化数据到Graph的转换
- **不懂图查询**：通过Text2GQL完成自然语言对话生成图查询并返回结果
- **消除LLM幻觉**：结合知识图谱与大模型提升问答质量
- **领域智能专家**：Chain of Exploration, 串联多种技术工具

## KG enablers

- KG Constructor
- QueryCraft
- Graph Architect

## Intelligence Accessors

- QueryCraft
- Graph RAG
- Chain of Exploration

## Agent Infra & Tool

- Accessor Connectors
- Graph Memory
- Meta DAG Keeper

Infra Components

Solution Components

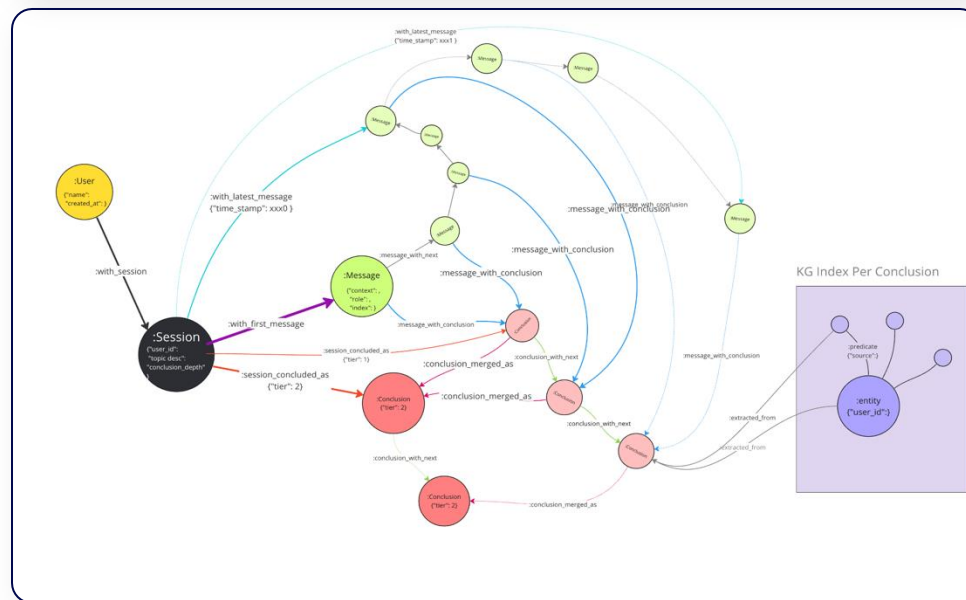
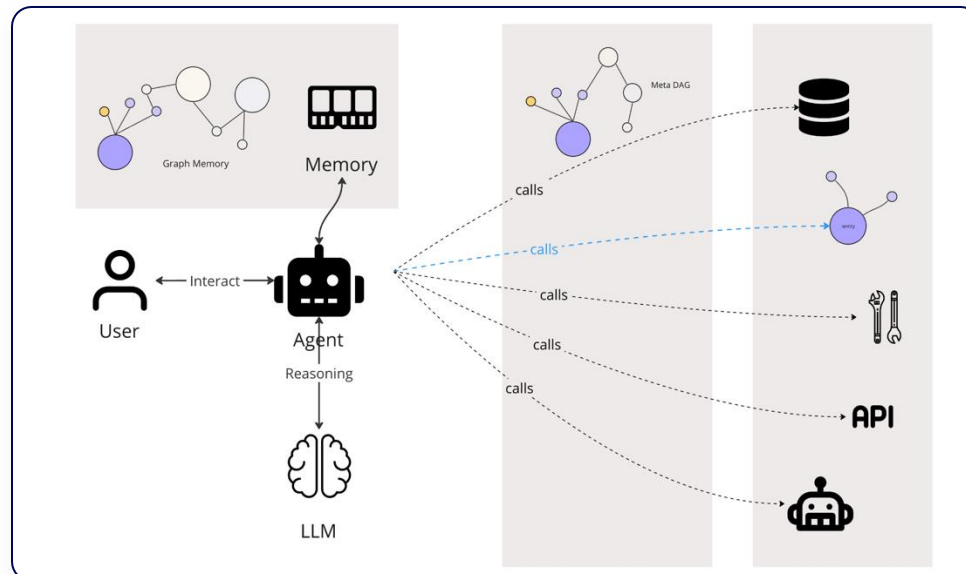
## Enhanced Models & Toolchain

- InsightPlus KG Models
- FinTech Explorer Model
- HealthGraph Model
- Domain Tuner Kit
- Toolchain and Workflow to train Industry/Custom Model

## Knowledge Services/Products

- API
- Accessors
- API/Daemon
- Knowledge Chat
- Knowledge Studio

Model Components



# 2 NebulaGraph Community



已获得  
**11k** stars  
**1.2k** forks

 **nebula** Public

A distributed, fast open-source graph database featuring horizontal scalability and high availability

 C++  11k  1.2k

**40%+**  
MAU

- 论坛用户数为 3k;
- 微信群有 2k 的活跃成员



**30 k+**  
post

- 120 万的 PV 和 4k+ 的 Topic;



**2k+**

开发者总数

- GitHub、论坛、微信群



**NebulaGraph**



**~10k**  
周下载量

- 可以在 GitHub、Docker Hub、官网三个主要渠道下载安装包

# 海内外开发者分布





**20+**  
高校合作

- 高校产学研合作
- 如：复旦大学、University of North Carolina Asheville 等



**200+**  
社区合作

- 开源社区合作遍布海内外
- 如：LlamaIndex、CAMEL AI 等



**2000+**  
企业用户

- 分布在互联网、金融、政府、运营商、制造业、零售业、公用事业等各个行业
- 20+ 个千亿点边规模



# NebulaGraph 社区活动

## 大版本 捉虫

将使用过程中遇到的问题  
报给社区，一起修复 bug  
完善产品

## nMeetup 城市站

国内外大厂技术团队最佳  
NebulaGraph 实践

## Nebula Hackathon

用创造力看见不一样的图  
实践

## 高校合作

整合学术资源与实践经验，  
促进创新、提高技术应用  
能力，并培养学生的实际  
技能与职业素养

# 近期活动, NebulaGraph nMeetup 厦门站

NebulaGraph 走进企业、高校、合作伙伴, 积极推进图数据库和 GraphRAG 的知识普及、经验交流、创新合作



## 图数据库+AI

NebulaGraph 社区 Meetup 第一期 厦门站

2024年8月17日 13:30-16:40

厦门市湖里区新丰路176-180号火炬荟智空间A区二楼路演厅

### 核心亮点

GraphRAG

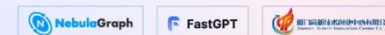
GQL 生态

AI 应用

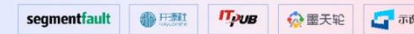
### 活动日程

- 13:30-13:40 ● 开场演讲  
内伯科技 | 周长威
- 13:40-14:10 ● 当传统文化遇见 GraphRAG  
厦门大学平潭研究院 | 杨知铮
- 14:10-14:40 ● AI 在客服中心录音质检的应用  
厦门巅峰科技 | 涂标松
- 14:40-15:10 ● 茶歇
- 15:10-15:30 ● 关于图数据库驱动程序的现状与倡议  
笃笃科技 | 叶伟城
- 15:30-16:00 ● 低代码赋能 AI 应用开发: FastGPT 高级编排实战  
环界云计算 | 王子俊
- 16:00-16:40 ● 面向 ISO-GQL  
杭州悦数 | 方扬

主办方



合作方





# 近期活动, NebulaGraph nMeetup 北京站

NebulaGraph 走进企业、高校、合作伙伴, 积极推进图数据库和 GraphRAG 的知识普及、经验交流、创新合作

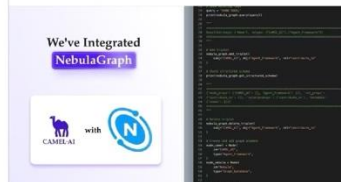


**Dinah Zhang**  
@dinah\_zhang

#NebulaGraph ❤️ #CamelAI  
For knowledge graph building #GraphRAG, we have added a new, quick start method, and a series of demos will be coming soon. Welcome to experience and share your feedback.  
Thanks to @ttokzzzzz, @VoidAsuka, ZackYule, raywhoelse for the contribution.  
翻译帖子

**CAMEL-AI.org** @CamelAIOrg · 1小时  
We've integrated @NebulaGraph in the CAMEL framework for knowledge graph building!

NebulaGraph is an open-source, distributed, scalable, lightning-fast graph database built for...



## GraphRAG meets AI

NebulaGraph 社区 Meetup 第二期 北京站

🕒 2024年9月21日 13:25-17:30

📍 北京市海淀区丹棱街5号微软大厦1号楼 故宫会议室

### 活动日程

- 13:25-13:30 **开场演讲**  
滑铁卢大学 | 张天翊
- 13:30-13:50 **金融领域大模型 GraphRAG 和图谱的结合应用**  
语势科技技术负责人 | 岳洋
- 13:50-14:10 **NebulaGraph 在博睿数据的应用实践**  
北京博睿宏远数据大数据研发 | 姚成耀
- 14:10-14:50 **GraphRAG in NebulaGraph**  
NebulaGraph GenAI Leader, Microsoft MVP | 吉思为
- 14:50-15:30 **Agents with GraphRAG**  
Eigent AI Founding Engineer | 范文栋
- 15:30-15:45 **茶歇&合照**
- 15:45-16:20 **AI Agent 企业落地实践与案例拆解**  
53AI 创始人, LangGPT 联合创始人 | 杨秀贤
- 16:20-16:50 **优化 GraphRAG 的二三事**  
微软高级技术布道师 | 卢建辉
- 16:50-17:30 **FastGPT 让企业用 AI 更简单**  
FastGPT maintainer, Sealos maintainer | 杨道升

主办方



合作方





# 近期活动, NebulaGraph nMeetup 上海站

## NebulaGraph 走进企业、高校、合作伙伴, 积极推进图数据库和 GraphRAG 的知识普及、经验交流、创新合作



## Database + AI Infra

NebulaGraph 社区 Meetup 第三期 上海站

2024年10月26日 13:30-18:10

携程集团总部 | 上海市长宁区金钟路968号12号楼11F02维多利亚会议室

### 活动日程

- 13:20-13:30 主持暖场&携程集团欢迎致辞
- 13:30-14:10 业务增益 32%, NebulaGraph 在携程风控场景下的落地实践  
携程集团 风控系统资深工程师 | 周伟斌
- 14:10-14:50 携程 NebulaGraph 图数据库平台建设  
携程集团 高级云原生研发工程师 | 郑皓月
- 14:50-15:10 NebulaGraph Flink Connector 在实时 ETL 中的应用实践  
新奥新智 实时计算平台开发 | 祝亚运
- 15:10-15:40 圆桌: 开源赋能, 高校学生与开发者的成长  
尹航, LFAPAC 开源布道师, NebulaGraph Ambassador (主持人)  
古墨为, NebulaGraph GenAI Leader, Microsoft MVP  
车洋, CNCF 项目 Fluid 联合创始人, 阿里云容器服务高级技术专家  
肖卓雅, Apache OpenDAL PMC Member  
Steven Lynn, Dify Technical Writer  
余涛, 复旦大学研二学生, 研究方向为图基础算子加速
- 15:40-16:00 茶歇&合照
- 16:00-16:30 向量数据库—AI 时代的基础设施  
Zilliz 资深开源布道师 | 李成龙
- 16:30-17:00 关于 GraphRAG 的思考, 进展和展望  
NebulaGraph GenAI engineer | 余毅航
- 17:00-17:30 ChatDBA: 数据库根因分析智能助手  
爱可生 AI 创新事业部负责人 | 苏鹏
- 17:30-18:00 FastGPT 企业级 AI Agent 落地实践  
FastGPT maintainer | 朱浩翥

开放麦: Think in Graph 讨论 (18:00-18:30)  
参与就送社区周边





# 近期活动，NebulaGraph nMeetup 深圳站

## NebulaGraph 走进企业、高校、合作伙伴，积极推进图数据库和 GraphRAG 的知识普及、经验交流、创新合作



**Graph Meets AI**  
NebulaGraph 社区 Meetup 第四期 深圳站

2024年11月30日 13:30-18:00  
深圳市南山区粤海街道科技园二期19楼1901大教室

### 活动日程

- 13:30-14:10 图数据库 NebulaGraph 在奇富科技应用实践  
奇富科技 风控架构师 | 李权
- 14:10-14:50 覆盖云原生 DevOps 全流程的 NebulaGraph 图数据库平台建设  
华为 高级工程师 | 支关记
- 14:50-15:20 圆桌：AI 对开源社区和开发者的影响  
嘉宾：Shopee Expert Engineer (主持人) 李权, 奇富科技风控架构师 张晋友, 广东智康人工智能应用研究院工业&社区 CTO, 微软 MVP 李东阳, 深圳开生态发展负责人, 开放社区并社区生态发展组组长 杨康博, LFAC 开源运营, 图Community 学术工业社区发起人 苏辉, CNCF Ambassador, 中国开源社区主持人, 腾讯 SDE
- 15:20-15:30 茶歇&合照
- 15:30-16:05 GraphRAG in NebulaGraph  
NebulaGraph GenAI 负责人 | 占志为
- 16:05-16:40 用 BentoML 部署你自己的机器学习应用  
BentoML 高级工程师 | 胡希
- 16:40-17:15 企业 AI 应用的提升之道：FastGPT 使用指南  
FastGPT 开发者布道师 | 王子俊
- 17:15-17:50 GraphRAG.NET 与 NebulaGraph  
世纪成云高级 .Net 开发工程师 | 甄家东

开放麦：Think in Graph 讨论 (18:00-18:30)  
参与社区周边

活动报名

主办方

NebulaGraph | 奇富科技

联合主办

英特尔大湾区科技创新中心  
Greater Bay Area Innovation Center

合作伙伴

FastGPT | BentoML | scalos | Broadview

合作社区

BentoML | scalos | Broadview | IT-PUB | milvus

# 3 高校合作计划



## 高校合作专区，合作价值

### 技术创新与融合

高校教师的研究与 NebulaGraph 社区的技术能力相结合，共同探索图数据库与知识图谱的创新方向，推动技术成果的转化与应用。

### 人才培养与发展

NebulaGraph 开源社区与高校合作，通过图数据库的学习与实践，提升人才培养的有效性，培养具备图数据库与知识图谱领域专业知识的人才。

### 教学方法与创新

NebulaGraph 社区支持高校教师将图数据库与知识图谱知识整合进教学，革新教学方法与实践，以培养具备实操能力的专业人才。

### 开源文化与合作

协助和鼓励高校人才参与开源项目，为图数据库技术的发展贡献力量，共同提升领域的竞争力及开源社区的良性发展。

# 高校合作专区，合作权益

## 产教融合与课程开发

企业与高校联合开发图数据库和知识图谱系列课程，供学生学习，培养专业人

## 高校沙龙和讲座支持

高校可申请 NebulaGraph 专家的支持，开展技术沙龙讲座，分享前沿技术知识创新与融合

## 职业发展通道

高校学生可申请参与 NebulaGraph 暑期实习计划与人才培养计划，加速职业发展

## 开发者大赛

高校教师指导学生报名参与 NebulaGraph 开发者大赛，提升学生的实践能力与实战经验

## 开源实践活动

高校学生可参与 NebulaGraph 社区的开源实践，如版本内测和日常实践任务，增强实际操作能力技术创新与融合

## 技术峰会参与

高校教师和优秀学生可受邀参加 NebulaGraph 技术峰会，共同探讨技术创新与行业应用

## 校园大使招募中

**校园大使**：作为 NebulaGraph 品牌的高校代言人

**义务**：协助策划并组织校园活动，传播图数据库、知识图谱技术，与 NebulaGraph 一起共建专业知识体系

**收获**：社区提供全方位官方资源支持，助力校园大使在校园中脱颖而出；有机会获得实习绿色通道，为未来职业发展增添竞争力。



**扫码报名**

# 感谢聆听

-  <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

# Graph DAG + 教育

## 圆桌主题：高校学生如何更好的参与开源社区



**尹珉 (主持人)**

LFAPAC 开源布道师  
NebulaGraph Ambassador



**颜嘉麒**

南京大学信息管理学院教授



**Sherman**

NebulaGraph PMC



**李梦**

中科院软件所  
开源之夏运营负责人



**古思为**

NebulaGraph  
GenAI Leader

# NebulaGraph 图数据库 和 NebulaGraph RAG 介绍

分享人：古思为

PPT 地址（部分视频和图片显示需要翻墙）：[https://siwei.io/talks/nju\\_2024/](https://siwei.io/talks/nju_2024/)



# 茶歇时间 (10分钟)



高校行·南京大学站活动群

# 如何参与开源：NebulaGraph 社区开发者经验分享

分享人：宋昊成



开源项目的介绍



开源收获



如何参与开源



Part **1**

# 参与开源

## ➤ 开源之夏

“开源之夏（英文简称 OSPP）”是中国科学院软件研究所“开源软件供应链点亮计划”指导下的系列暑期活动，旨在鼓励在校学生积极参与开源软件的开发维护，培养和发掘更多优秀的开发者，促进优秀开源软件社区的蓬勃发展，助力开源软件供应链建设。

### 开源之夏 活动流程

\*默认时区为北京时间 (UTC+8)



## 项目申请

8	234db0007	升级完善 KataContainers 3.x 日志系统	龙蜥社区 OpenAnolis	进阶/Advanced	中文&English	Kata Containers 3.x 日志配置与日志输出 Proposal , 并实现按 Conguration 调配各组件日志输出 Kata Containers 3.x runtime-rs 日志输出方式改善 , 规避 Logger 多次创建销毁 Kata Containers 3.x runtime-rs 兼容其他 VMM 日志输出实现	编程语言: Rust	要了解过于深入 难度过高	不行
9	236920008	实现 mybatis-plus Db类、Wrapper相关 kotlin方法	苞米豆	基础	中文	针对gitee pr 223新建的Db类以及Wrapper相关, 实现kotlin相关方法 Db.ktQuery();实现 Db.ktUpdate();实现 KtQueryChainWrapper将BaseMapper参数改为可选的 相关的单元测试 相关的中文注释完善	阅读过Mybatis-Plus 源码或 达到贡献能力者, 且熟悉kotlin	仔细了解不难 但是没接触过 从头要学习东西	行
10	236920009	dynamic-datasource-spring-boot-starter 分模块 兼容 springboot2、springboot3	苞米豆	基础	中文	实现分模块兼容springboot2、springboot3 Springboot3已发布,但是其更改了httpSelvet等相关的包名, 需要对其进行兼容 多数数据源内部内置了基于请求头,请求Header等切换方式. 当用户 使用了相关方法但运行在springboot3的项目就会报错,找不到相关的class,需要解决这个问题	有兼容springboot2、springboot3相关经验		不行
11	230900010	xmake-repo 仓库包制作	xmake	基础	中文	完成 nanobind, sleepy-discord, libgda, gtk+3, gtk4 包制作并入库 xmake-repo 至少支持 linux 平台 (例如 gtk+3, gtk4) , 部分跨平台的库, 可以支持上 macos 和 windows 平台	熟悉和了解 xmake 基本使用和 仓库包制作流程 熟悉 cmake/meson 等其他构建工具 熟悉 linux/windows 下开发 熟悉 C/C++ 开发, 了解基础的 编译链接机制	库涉及的依赖比较复杂的工作量, 中间可能会遇到各种编译问题需要解决, 也具有一定的挑战性。	行
12	2357b0012	MQTT Broker消息持久化	smartboot	进阶	中文	完成桥接插件代码的开发 在MQTT Broker的启动进程中成功加载插件 桥接插件成功将消息写入redis, 并通过redis客户端工具 确认消息有效	掌握smart-mqtt插件的开发、加载技巧 运用Docker快速搭建redis环境	有一定技术门槛 可以做, 但是会比较难 但是申请的人可能不多	行

## 充分准备

- **仔细分析:** 项目涉及的技术栈, 编写难度, 自己的技术能力是否匹配, 底线是保证项目完成
- **有效沟通:** 跟导师交流顺畅能极大提高项目的成功率
- **写好申请书:** 项目涉及的技术细节、技术方案、项目时间规划

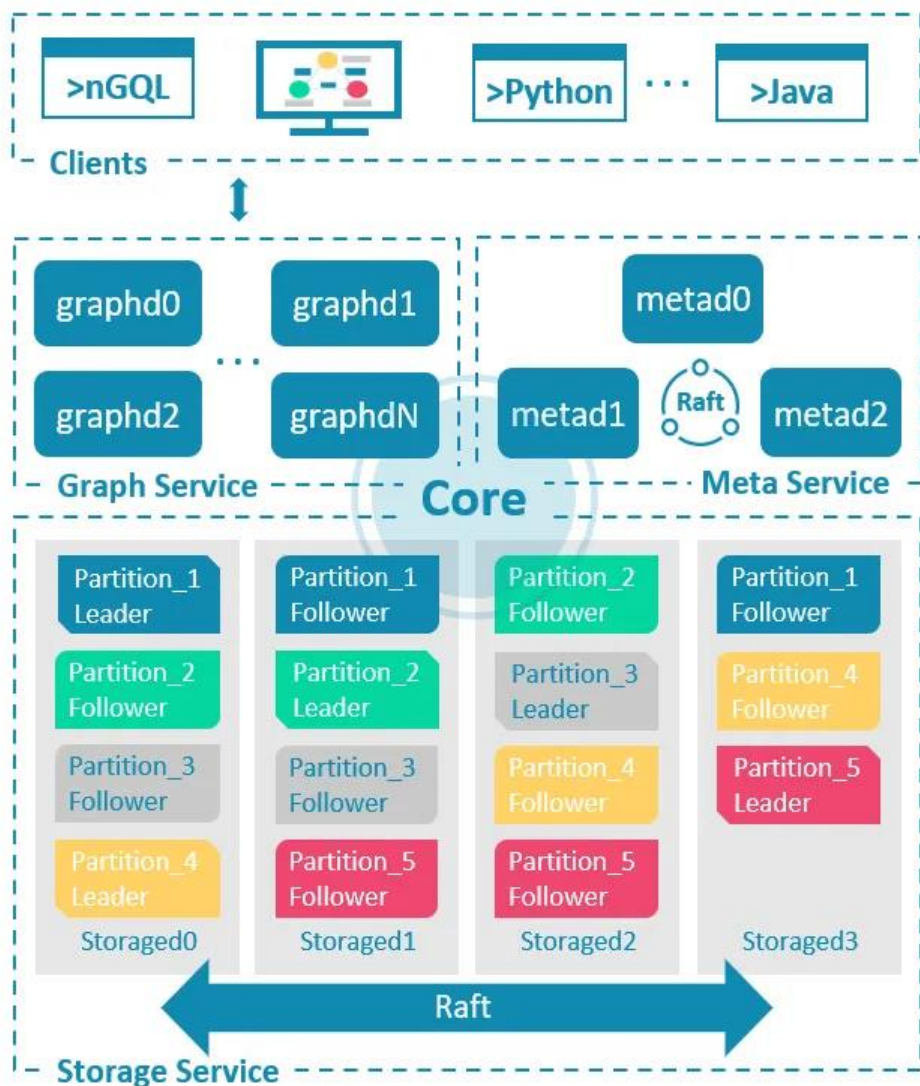




Part 2

# 开源项目介绍

## 项目介绍



**NebulaGraph:** NebulaGraph 是一个高性能的分布式图数据库，采用存储与计算分离的架构，由 Graph、Meta 和 Storage 服务构成。

**NebulaGraph-rust 实现点边扫描接口:** NebulaGraph-rust 仅支持部分 GraphClient 接口，需要完成对 Storage 服务的访问，提升项目在 Rust 环境中的完整性和功能性。

### 实现功能

- 实现Nebula-Rust客户端的建立，包括建立连接、相关命令构建、命令发送和接收
- 基于Graph-Service完成根据nGQL语句查询并返回查询结果
- 完成ScanVertex和ScanEdge接口，返回点和边的扫描结果。（提供两种反序列化方案）

## ➤ 实际演示

**扫描点和边结果进行返回：** 连接到 192.168.10.21:9559 的 NebulaGraph 数据库，扫描 **basketballplayer** 空间中标签为 **player** 的顶点和类型为 **serve** 的边。

**输入命令：** `cargo run -p nebula-demo-tokio --bin nebula_demo_scan 192.168.10.21 9559 basketballplayer player basketballplayer serve`

```
Part ID: 1, Leader: 192.168.10.21:9779
(player105:player{name: Danny Green, age: 31})
(player109:player{name: Tiago Splitter, age: 34})
(player111:player{name: David West, age: 38})
(player118:player{name: Russell Westbrook, age: 30})
(player143:player{name: Kristaps Porzingis, age: 23})
Part ID: 2, Leader: 192.168.10.21:9779
(player114:player{name: Tracy McGrady, age: 39})
(player150:player{name: Luka Doncic, age: 20})
Part ID: 3, Leader: 192.168.10.21:9779
(player103:player{name: Rudy Gay, age: 32})
(player113:player{name: Dejounte Murray, age: 29})
(player121:player{name: Chris Paul, age: 33})
(player128:player{name: Carmelo Anthony, age: 34})
(player130:player{name: Joel Embiid, age: 25})
(player136:player{name: Steve Nash, age: 45})
```

ScanVertexResult(以前三个分区为例)

```
Part ID: 1, Leader: 192.168.10.21:9779
(player105)-[:serve@0{start_year: 2010, end_year: 2018}]->(team204)
(player105)-[:serve@0{start_year: 2009, end_year: 2010}]->(team216)
(player105)-[:serve@0{start_year: 2018, end_year: 2019}]->(team218)
(player109)-[:serve@0{start_year: 2010, end_year: 2015}]->(team204)
(player109)-[:serve@0{start_year: 2017, end_year: 2017}]->(team219)
(player109)-[:serve@0{start_year: 2015, end_year: 2017}]->(team222)
(player111)-[:serve@0{start_year: 2016, end_year: 2018}]->(team200)
(player111)-[:serve@0{start_year: 2015, end_year: 2016}]->(team204)
(player111)-[:serve@0{start_year: 2003, end_year: 2011}]->(team215)
(player111)-[:serve@0{start_year: 2011, end_year: 2015}]->(team220)
(player118)-[:serve@0{start_year: 2008, end_year: 2019}]->(team205)
(player143)-[:serve@0{start_year: 2019, end_year: 2020}]->(team213)
(player143)-[:serve@0{start_year: 2015, end_year: 2019}]->(team223)
Part ID: 2, Leader: 192.168.10.21:9779
(player114)-[:serve@0{start_year: 2004, end_year: 2010}]->(team202)
(player114)-[:serve@0{start_year: 2013, end_year: 2013}]->(team204)
(player114)-[:serve@0{start_year: 1997, end_year: 2000}]->(team218)
(player114)-[:serve@0{start_year: 2000, end_year: 2004}]->(team226)
(player150)-[:serve@0{start_year: 2018, end_year: 2019}]->(team213)
```

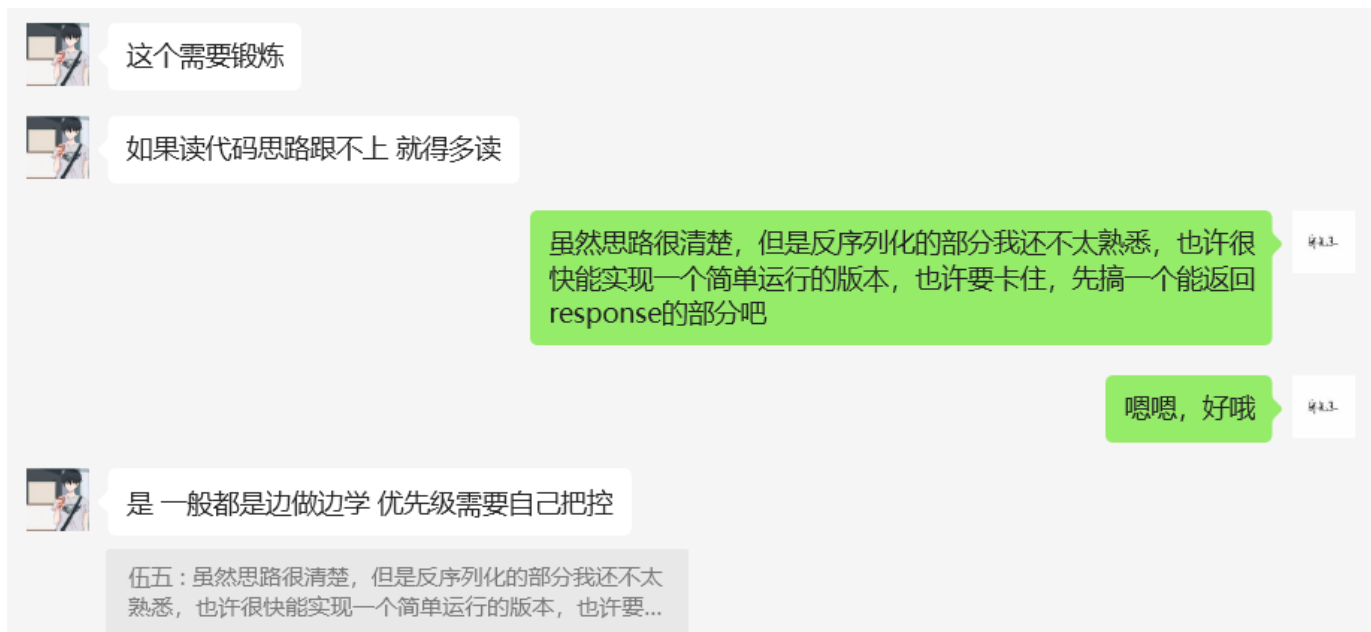
ScanEdgeResult(以前两个分区为例)



## 项目开发经验

**完成比做的完美更重要：**采取一种渐进的方法。首先从实现一个最基本的功能开始，确保它能够正常工作，然后在此基础上逐步扩展，最终形成想要的 mod（模块）或者 crate（包）

**及时突破瓶颈：**当项目进度停滞时，需要积极的推动，通过与导师沟通、网上交流、大模型协助等方式，不能长时间卡在一个进度



积极沟通



米泽双 github:Milittle



Part 3

# 开源收获

➤ 开源收获

收获：足够的自信、社区的周边、社区的友谊、认识各位技术大牛、优秀的过往经历



Contributor证书



明信片



结项证书



求职的帮助

# 制造差异点：开源经历让你有别于其他应届生，起到锦上添花的作用



Hi, 亲爱的宋昊成同学:

鉴于您在面试中的优秀表现, 我们很高兴发出这封录用通知书给您, 真诚地欢迎您加入这个积极、向上、阳光、激情洋溢的滴滴 family!

滴滴

campus 下午 06:35 ...  
发至 michael@mail、campus v

亲爱的同学:

您好!

我们非常高兴地通知您: 经过层层甄选, 您凭借出色的表现, 获得京东2025年度应届毕业生录用资格。加入京东, 以京东的速度启航梦想!

京东



快手

# 感谢聆听

-  <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

# Nebula Graph 在企查查的应用 实践

张宝利

# 背景

企查查是企查查科技股份有限公司旗下的一款企业信用查询工具 <https://www.qcc.com>



# 数据维度

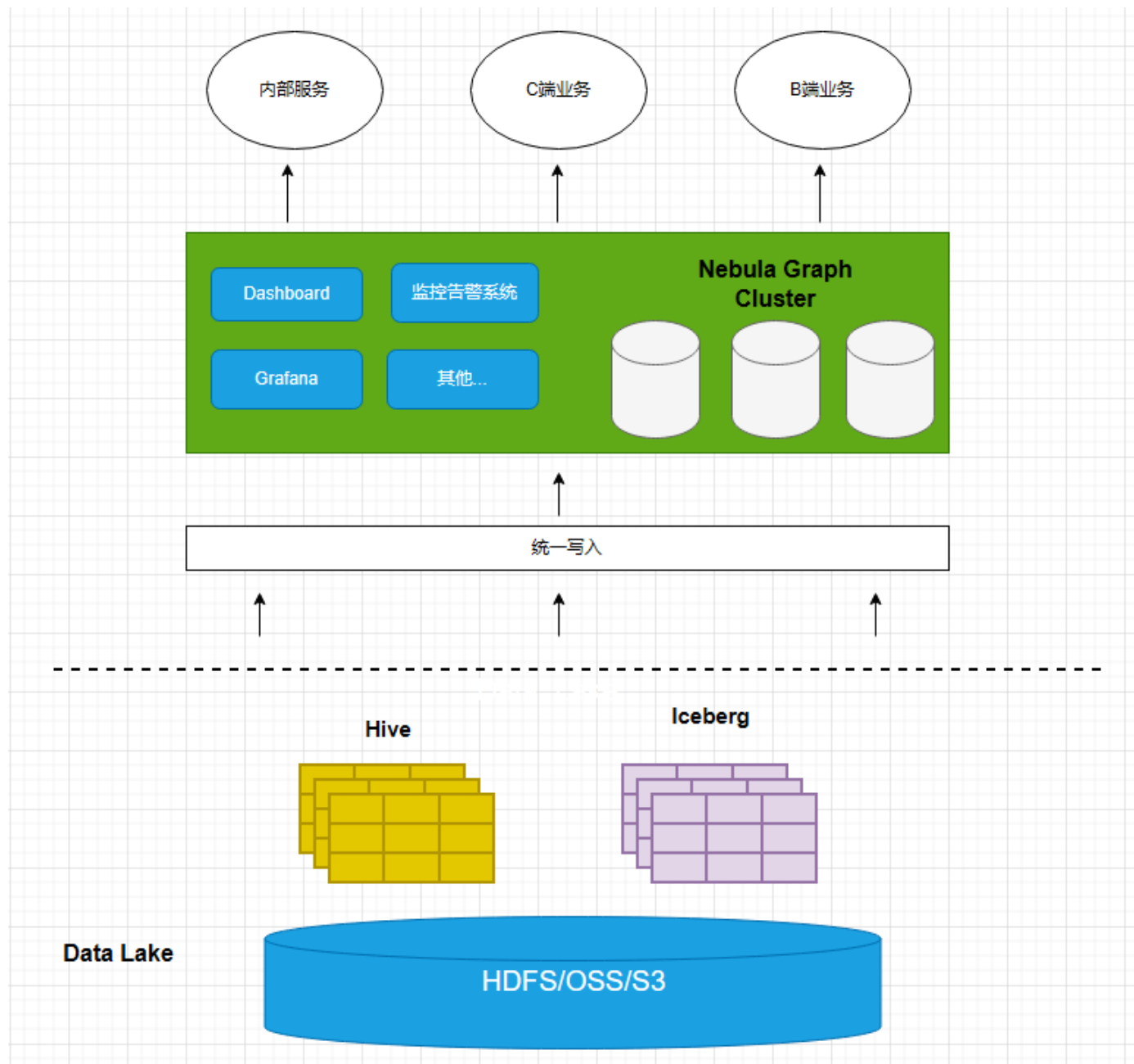
- 股东信息 (对外投资)
- 法人、历史法人
- 主要人员、历史主要人员
- 客户
- 供应商
- 分支机构
- 受益所有人
- 实际控制人
- ...





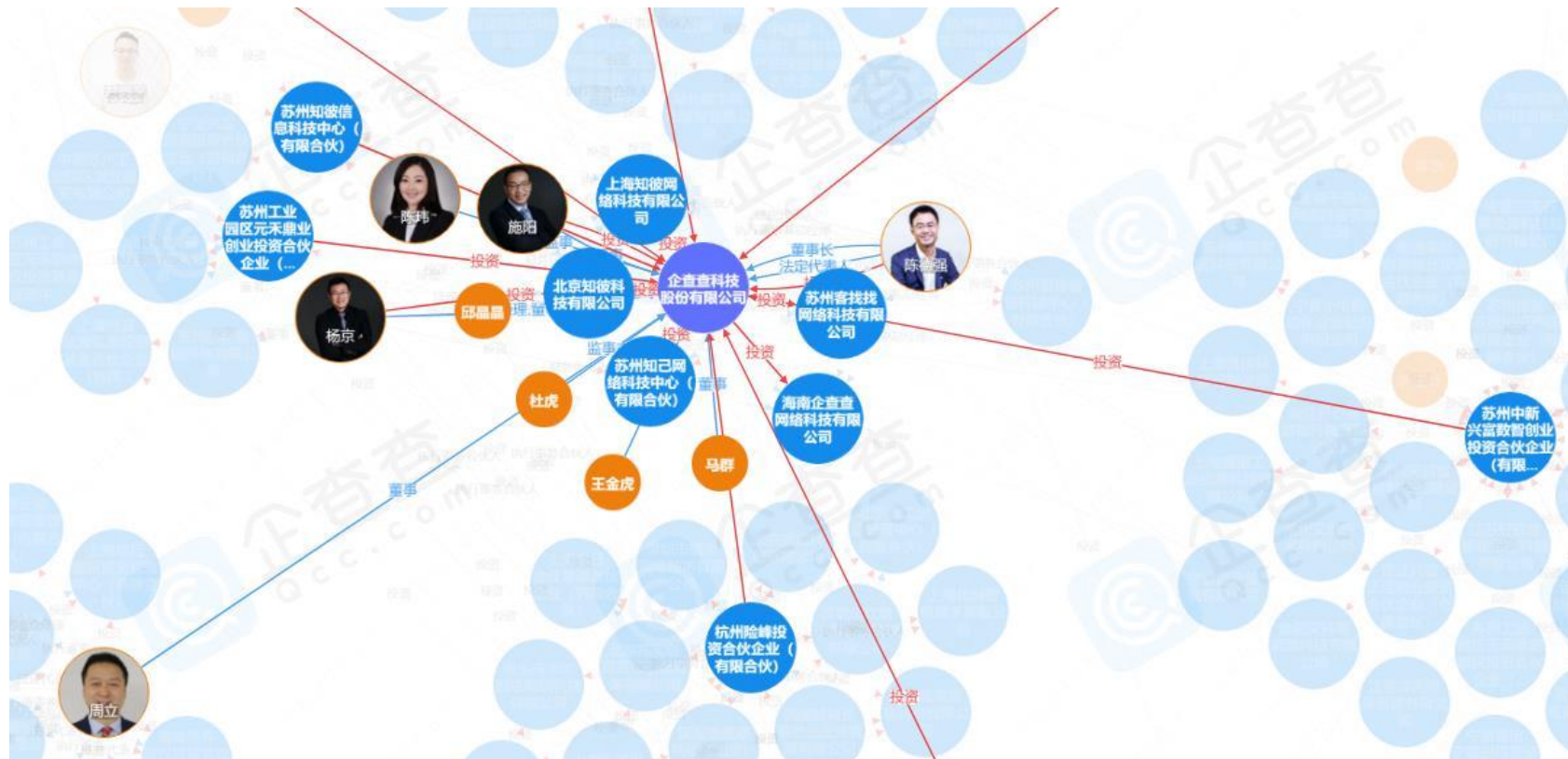
# 架构介绍

- 产品选型 neo4j -> nebula
  - 单机 -> 集群
  - T+1 -> 实时
  - 数据同步不间断
- 写路径
  - 全量 Spark 读hive 写入
  - 增量 Kafka batch + 时间戳排序后写入
- 读路径
  - SLB 负载均衡



## 业务介绍 — 关系图谱

- 获取主体的基础工商信息
  - 默认2 hops
- GET SUBGRAPH WITH PROP 2 STEPS FROM "A" BOTH Invest, Employ, Legal YIELD VERTICES as v, Edges as e



## 业务介绍 — 关系图谱

- 关系太多如何展示
  - 根据关系优先级, 实体优先级进行过滤
- 超级节点如何处理
  - 计算哪些实体是超级节点, 打上标签
  - 如果主体是超级节点只查询一跳, 否则根据业务需要选择过滤掉超级节点数据
  - `GET SUBGRAPH WITH PROP 2 STEPS FROM "A" BOTH Invest, Employ, Legal where $.Company.valid != true and $.Person.valid != true YIELD VERTICES as v, Edges as e`

# 业务介绍 — 风险排查

排查风险帮助用户公司“了解你的合作方”

- 关联方（投资公司，控制公司，关联方的关联方
- 裁判文书
- 经营异常
- 严重违法
- 失信被执行人
- 税收违法
- 行政处罚
- 开庭公告
- 电话，地址，邮箱
- ...

## 3.1、风险排查结果


风险等级：**【高风险】**

原因如下：

- 目标企业命中 **警示** 类风险项 **内部黑名单** 类型，包括：目标企业 **被列入内部黑名单**
- 目标企业命中 **关注** 类风险项 **法律风险** 类型，包括：目标企业 **重大纠纷、不正当竞争纠纷**
- 目标企业命中 **关注** 类风险项 **交叉重叠关系排查** 类型，包括：目标企业 **与第三方列表企业存在投资任职关联、与第三方列表企业存在潜在关联/利益关联方**

企查查科技股份有限公司

3 / 15

 企查查 第三方风险排查

联系电话：400-088-8275

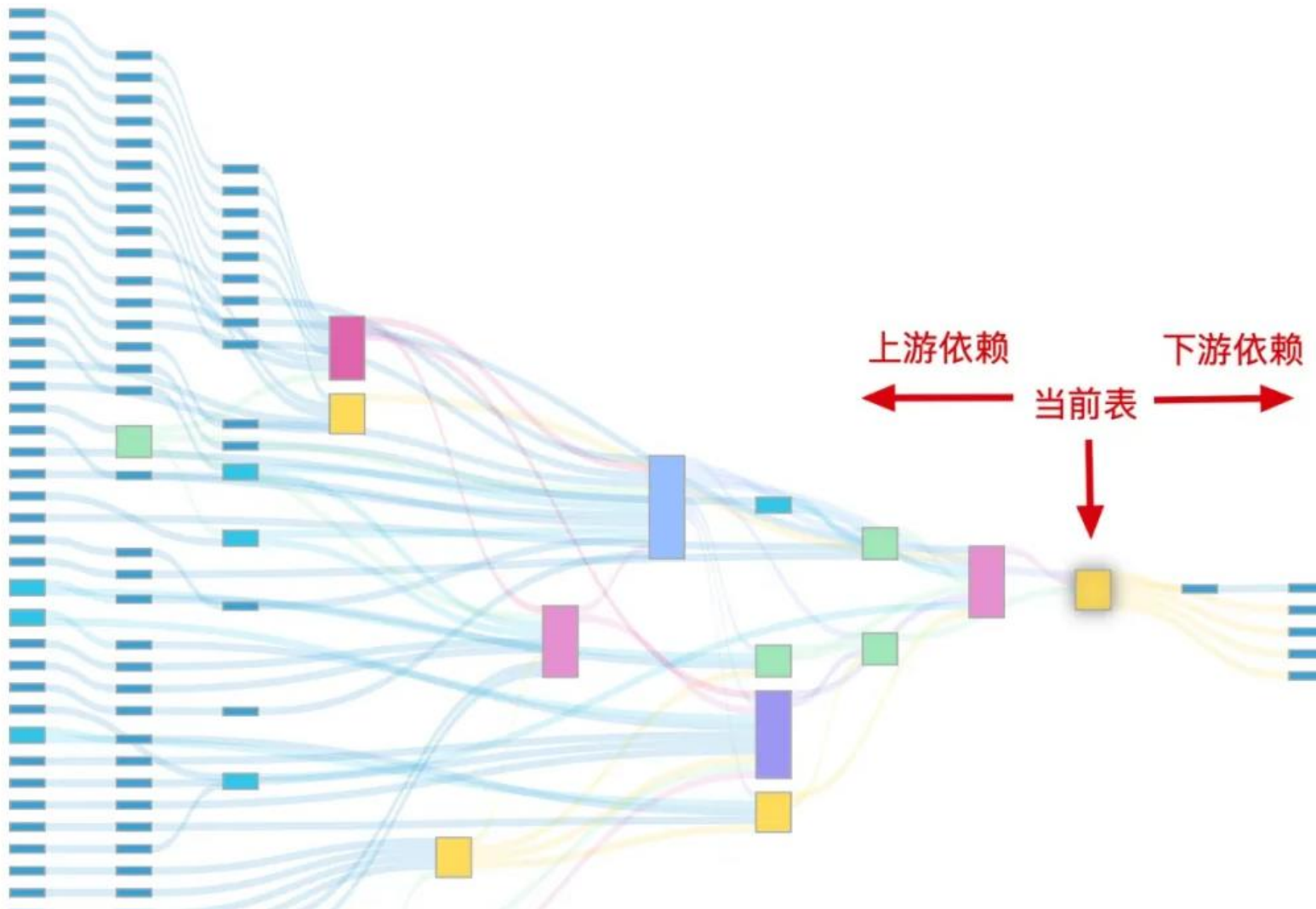
- 目标企业命中 **关注** 类风险项 **内部黑名单** 类型，包括：目标企业 **与内部黑名单企业存在投资任职关联、与内部黑名单企业存在潜在关联/利益关联方**
- 目标企业命中 **关注** 类风险项 **潜在利益冲突** 类型，包括：目标企业 **疑似潜在利益冲突**
- 目标企业【企查查】得分 **1182分**，信用表现 **卓越**，当前处于良性循环状态，**违约风险很低**

信用评分：



# 业务介绍 — 血缘关系

- 业务线
- Task
- Table
  - 字段
- Topic
- Group
- IP





## 常见优化

- "disable\_auto\_compactions": "true "
  - 每天增量3kw +
  - 业务高峰期CPU, IO 拉满
  - 提速spark writer 提速3x
- 定时job full compaction
  - 提高查询效率
- 写入, 减小写入压力
  - 增量控制好批次, batch 内相同类型去重
  - 设置TTL + 定时全量校正数据

# 常见优化

- 超级节点
  - max\_edge\_returned\_per\_vertex
  - 对实体打标签, 查询时额外处理, 比如过滤掉
  - 对实体进行拆分, 仅限个别业务
    - EntityA -> EntityA1, EntityA2, EntityA3
- 业务隔离
  - SLB Graph 隔离, 避免内存毛刺互相影响
- 其他治理
  - 数据治理
  - 稳定性治理

# 谢谢大家

-  <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

# 基于大语言模型的中医古籍糖尿病食疗知识图谱 构建及数据挖掘研究



主讲人：洪海蓝



时间：2025

# 目录

CONTENTS

01 研究背景与意义

02 研究方法与过程

03 研究成果与应用

04 结论与展望



# 研究背景与意义

-  <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

# 大语言模型与知识图谱的技术优势

## 大语言模型的强大文本处理能力

大语言模型如glm-4具备强大的自然语言处理能力，能够自动从非结构化的中医古籍文本中提取出有关糖尿病食疗的实体、概念和关系，为知识图谱的构建提供技术支撑。

### 知识图谱在数据组织与分析中的作用

知识图谱通过图形化结构表示和存储知识，能够有效地存储和表示实体和关系，为人工智能系统提供知识支撑，揭示疾病与药物、基因、生活方式等多维度因素之间的复杂关系。

### 大模型与知识图谱结合的创新意义

将大语言模型与知识图谱结合，能够实现从文本到知识的高效转换，为中医古籍的现代化研究和应用提供新的工具和视角，推动中医药膳食疗法调养糖尿病的现代化和国际化。

# 研究方法过程

-  <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>

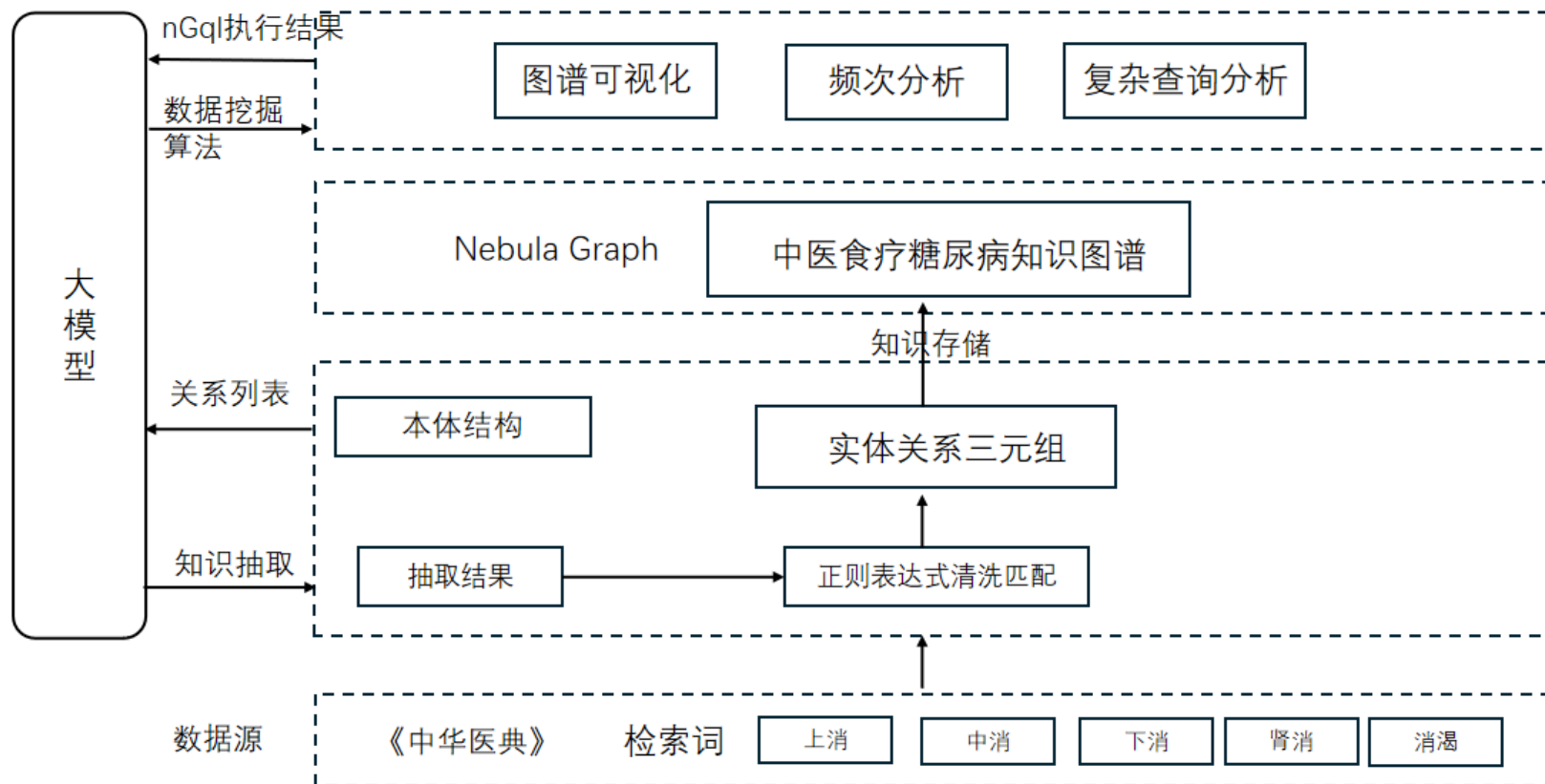


微信公众号



开源项目

# 中医古籍中药膳食疗方食疗糖尿病知识图谱的框架图



# 知识图谱的构建流程

01

## 数据源的整合与筛选

数据来源于《中华医典》等中医古籍，包括《黄帝内经》、《伤寒杂病论》、《千金方》、《圣济总录》等，以“消渴”等关键词进行检索，筛选出与糖尿病食疗相关的方剂，共939首方剂。



02

## 本体设计与实体关系确定

确定实体类型包括古籍、方剂、食材、药材等，以及实体之间的语义关系，如“古籍-方剂”、“方剂-食材”、“方剂-药材”等，构建出包含4种实体类型、5种语义关系的知识图谱本体结构。



03

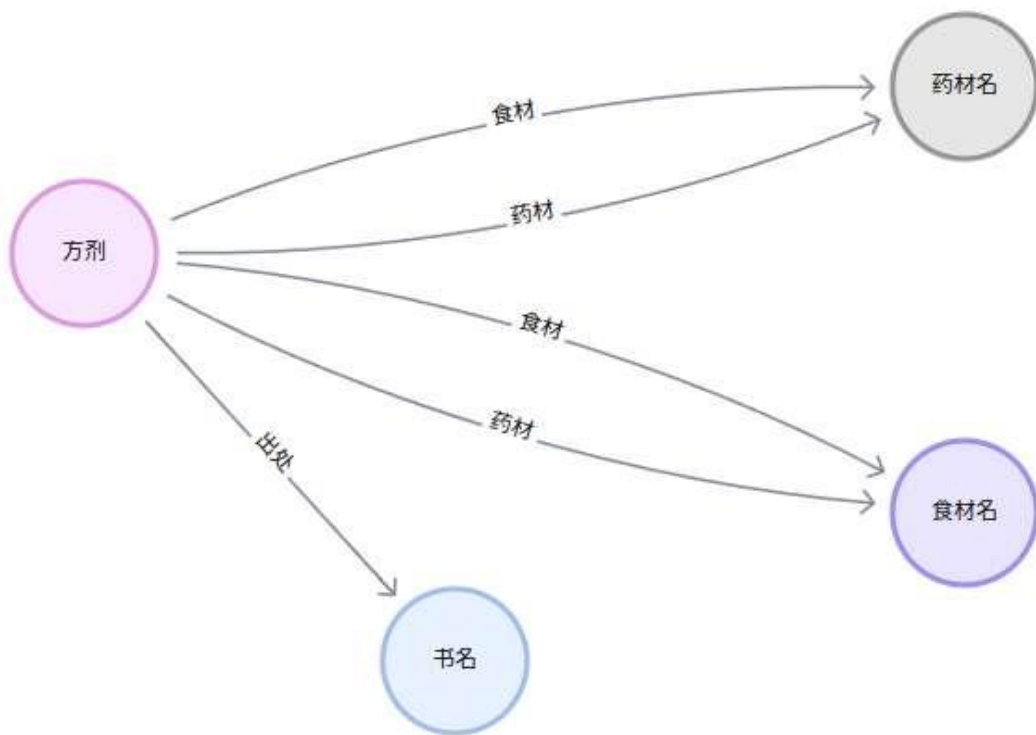
## 知识图谱的存储与管理

使用Nebula Graph图数据库存储和管理知识图谱数据，其强大的可视化图形工具Nebula Graph Studio提供了直观的Web界面，支持批量导入数据和探索图数据的功能。





# 中医古籍中药膳食疗方食疗糖尿病知识图谱schema



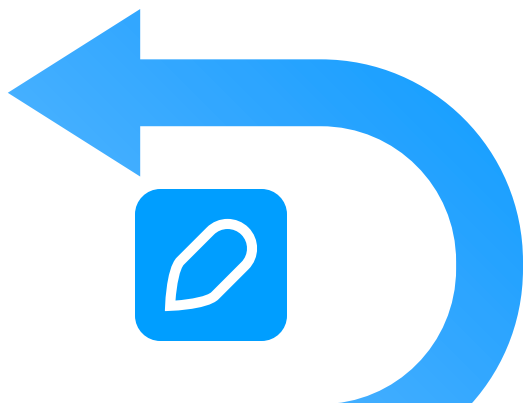
# 基于LLM的知识抽取算法

a. 通过模型开放的api\_key与大语言模型建立连接，同时对大语言模型进行prompt，比如，你是{一个中医古籍中药膳食疗方食疗糖尿病的关系抽取大模型,对古籍记载医案进行关系抽取,只抽取存在的关系}

b. For i=1 to N do

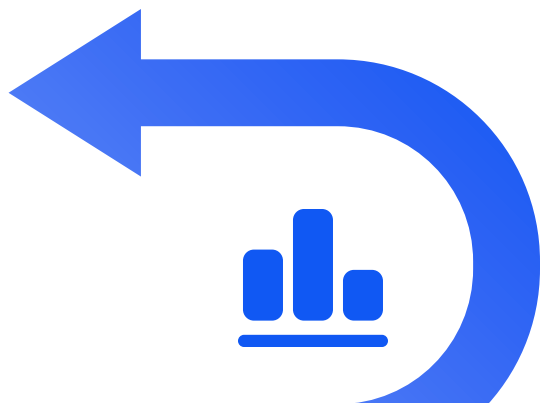
- ① select(D)//选择一条相关数据,生成相关的提问题。
- ② Prompt //给定关系列表: ['出处','方名','食材','药材']和相关的抽取案例。
- ③ 让大模型学习上述案例和关系列表对数据进行知识抽取，只出关系列表中的关系。如果不存在则不输出。
- ④ 大模型根据问题返回抽取的结果。
- ⑤ 利用正则表达式或者模式匹配对结果进行数据清洗，删除不必要的三元组。
- ⑥ End for

# 大语言模型的知识抽取



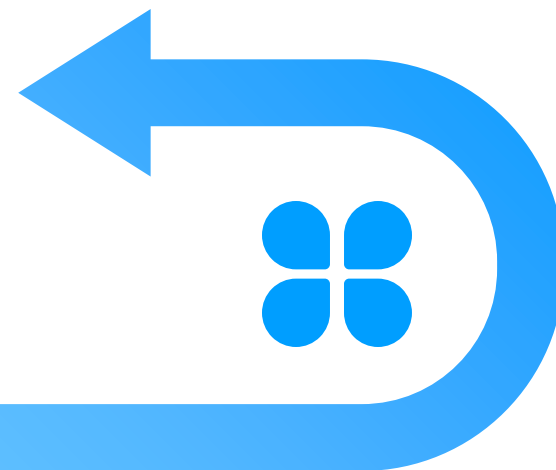
## 基于LLM的知识抽取算法

利用大语言模型的API接口，通过prompt引导模型进行知识抽取，如设置prompt为“你是中医古籍中药膳食疗方食疗糖尿病的关系抽取大模型”。对古籍中的方剂信息进行逐条抽取，提取出方剂名称、出处、食材、药材等实体及其关系，形成知识图谱的初步数据。



## 知识抽取的效果评估

实验结果表明，大语言模型glm-4在知识抽取方面表现优异，F1值达到了93%，P值和R值也表现良好，显示出其在自动化提取实体和关系方面的强大能力。



## 知识抽取过程中的挑战与优化

在知识抽取过程中，可能会遇到数据格式不统一、实体关系复杂等问题，需要通过优化算法和调整prompt等方式进行解决，以提高知识抽取的准确性和效率。

# 研究成果与应用

-  <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

# 知识图谱构建的成果

## 01



### 实体与关系的统计

构建的知识图谱包含939首方剂、1015种中药材、261种食材，形成了2416个实体和7817对三元组，全面展示了中医古籍中糖尿病食疗的丰富信息。

## 02



### 知识图谱的可视化展示

通过Nebula Graph Studio的可视化界面，用户可以直观地浏览知识图谱，深入了解各实体之间的关联，如人参白术汤图谱等。

## 03



### 知识图谱的应用价值

知识图谱为研究人员和临床医生提供了一个全面、系统的糖尿病食疗信息平台，有助于更好地理解糖尿病药膳食疗的组方规律，为患者提供个性化的膳食建议。



# 基于LLM的数据挖掘

- a. 通过模型开放的api\_key与大语言模型建立连接。
- b. 将中医古籍糖尿病食疗知识图谱的schema输入大模型：

```
{Create Space CREATE SPACE `diabetes`  
(partition_num = 10, replica_factor = 1, charset = utf8, collate = utf8_bin, vid_type =  
FIXED_STRING(100)); :sleep 20; USE `diabetes`; # Create Tag: CREATE TAG `书名` () ttl_duration = 0, ttl_col  
= ""; CREATE TAG `方剂` ( `方名` fixed_string(100) NULL) ttl_duration = 0, ttl_col = ""; CREATE TAG `药材名`  
() ttl_duration = 0, ttl_col = ""; CREATE TAG `食材名` () ttl_duration = 0, ttl_col = ""; # Create Edge: CREATE  
EDGE `出处` () ttl_duration = 0, ttl_col = ""; CREATE EDGE `药材` () ttl_duration = 0, ttl_col = ""; CREATE  
EDGE `食材` () ttl_duration = 0, ttl_col = ""};
```
- c. 给LLM提供nGQL的示范案例，请你根据用户需求生成对应的分析nGQL,严格按照关系格式，nGQL语句，分析，如MATCH p= (f:`方剂`)-[e:`药材`]->(c:`药材名`) return p。
- d. 用户提出问题，大语言模型根据用于需求自动化生成查询语句输出。
- e. 利用正则表达式清洗输出结果，获取nGQL语句，并进行查询，将查询结果进行图形可视化展示或者进行分析。

# 数据挖掘与分析

01

## 基于大模型的数据挖掘方法

用户向大语言模型提出具体问题，模型根据问题需求自动化生成对应的nGQL查询语句，实现从自然语言到图数据库查询语言的转换。

02

## 数据挖掘结果的可视化展示

执行查询语句后，将结果以图形化的方式展示，如通过表格和图形展示用药频次最高的前10味中药，使得分析结果更加直观和易于理解。

03

## 数据挖掘的应用场景与价值

通过数据挖掘，能够深入分析食材、药材与糖尿病之间的复杂关系，为研究人员和临床医生提供个性化的食疗建议，优化糖尿病患者的膳食方案。

# 数据挖掘的成果

## 01

### 简单查询的高效实现

在处理简单查询时，如查询用药频次最高的前10味中药，glm-4能够直接生成有效的nGQL语句，并以表格和图形化的形式展示结果，极大地提升了查询效率。

## 02

### 复杂查询的探索与挑战

当查询变得更加复杂时，例如查询记载方剂最多的古籍及其方剂的食材、药材等信息时，glm-4生成的查询语句需要进一步的调整和优化，揭示了其在处理复杂查询时存在的不足。

## 03

### 数据挖掘对临床实践的指导意义

数据挖掘结果能够为临床医生在制定糖尿病患者的个性化治疗方案时提供参考依据，如挖掘出某药材在不同方剂中的应用规律及其对糖尿病治疗的效果。

# 结论与展望

-  <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

# 研究的学术与实践意义



## 学术价值

本研究成功构建了中医古籍糖尿病食疗知识图谱，展示了大语言模型与知识图谱结合在数据挖掘领域的潜力，为中医传承与古籍医案挖掘提供了新的思路和方法，具有重要的学术价值。



## 实践意义

为全球糖尿病患者提供了更加科学、个性化的食疗建议，有助于优化糖尿病患者的膳食方案，提高糖尿病的治疗效果，具有重要的实践意义。



## 对中医药现代化的推动作用

促进了中医药膳食疗知识的传播和应用，提高了公众对糖尿病食养管理中作用的认识和理解，为中医药的现代化和国际化贡献力量。



# 研究结论

## 大模型在知识抽取中的优势



大语言模型在自动化提取实体和关系方面表现出色，能够显著提升知识图谱构建的效率和准确性，为中医古籍的现代化研究提供了强有力的技术支持。

## 数据挖掘对研究的深化作用



基于大模型的数据挖掘能够进一步挖掘中医古籍中蕴含的丰富知识，为研究人员提供更深入的洞见，推动糖尿病食疗研究的不断深入和发展。

## 知识图谱在糖尿病管理中的应用潜力



中医古籍糖尿病食疗知识图谱深入揭示了食材、药材与糖尿病之间的复杂关系，为糖尿病的预防、诊断和治疗提供了科学依据，具有广阔的应用前景。

# 未来研究方向

## 提高模型的复杂查询处理能力

针对大语言模型在处理复杂查询时的不足，未来的研究应重点关注如何提高模型的复杂查询处理能力，实现更自动化和智能化的数据挖掘过程。

## 知识图谱在临床应用中的探索

关注知识图谱在实际临床应用中的有效性和可行性，探索其在糖尿病患者个性化治疗方案制定、疗效评估等方面的应用，推动中医药膳食疗法在临床实践中的广泛应用。

## 中医药知识图谱的拓展与深化

进一步拓展和深化中医药知识图谱的研究，整合更多的中医药领域知识，构建更为全面和精准的知识图谱，为中医药的传承、创新和发展提供更为坚实的知识基础。

# 谢谢大家

-  <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

# 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

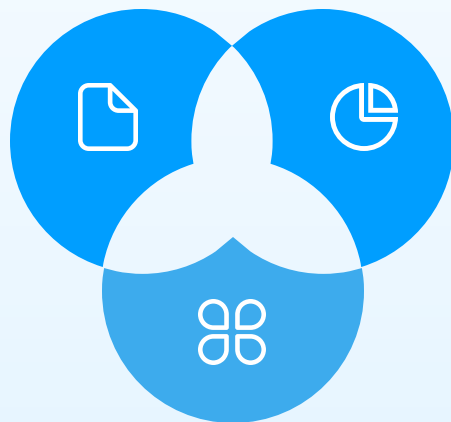
更小的交易损失量  
更低的用户打扰率



## 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

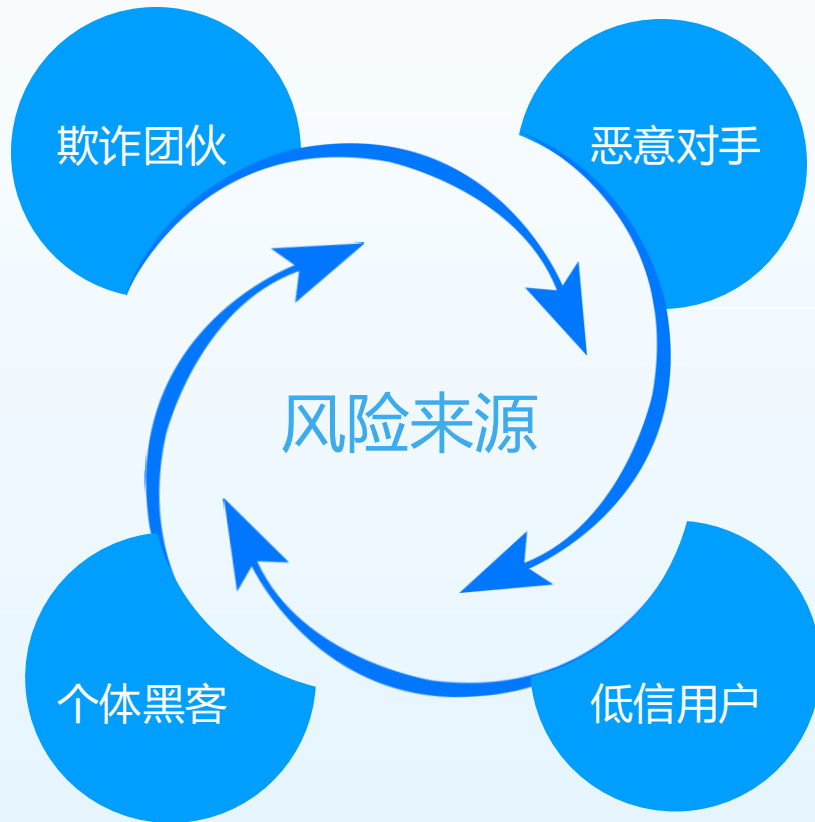
系统研发

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

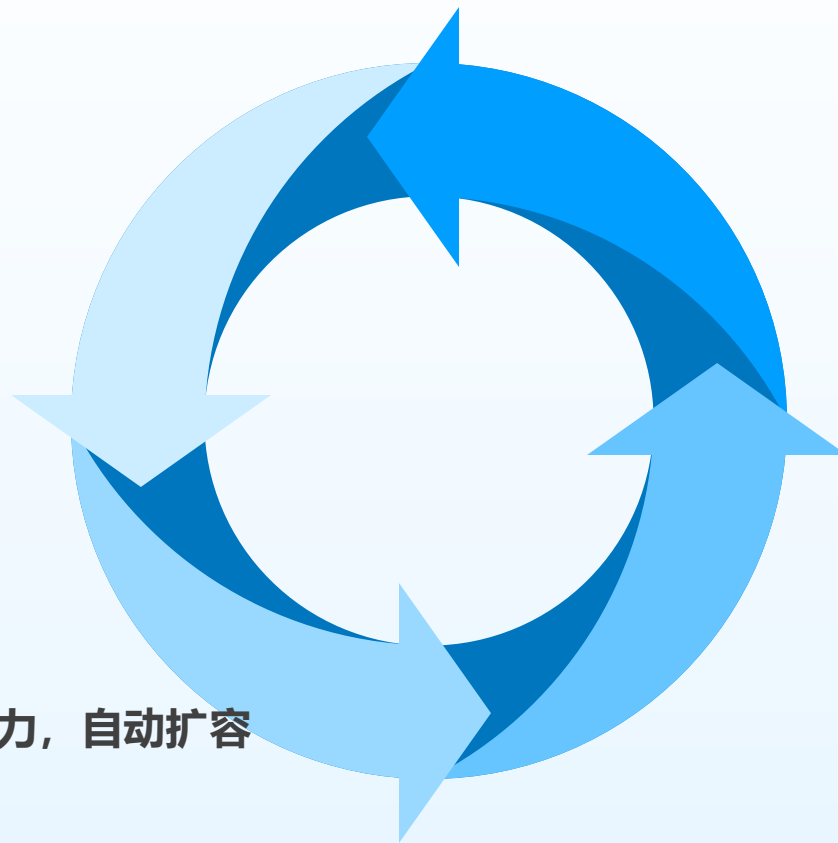
# 风控系统的保障

**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;

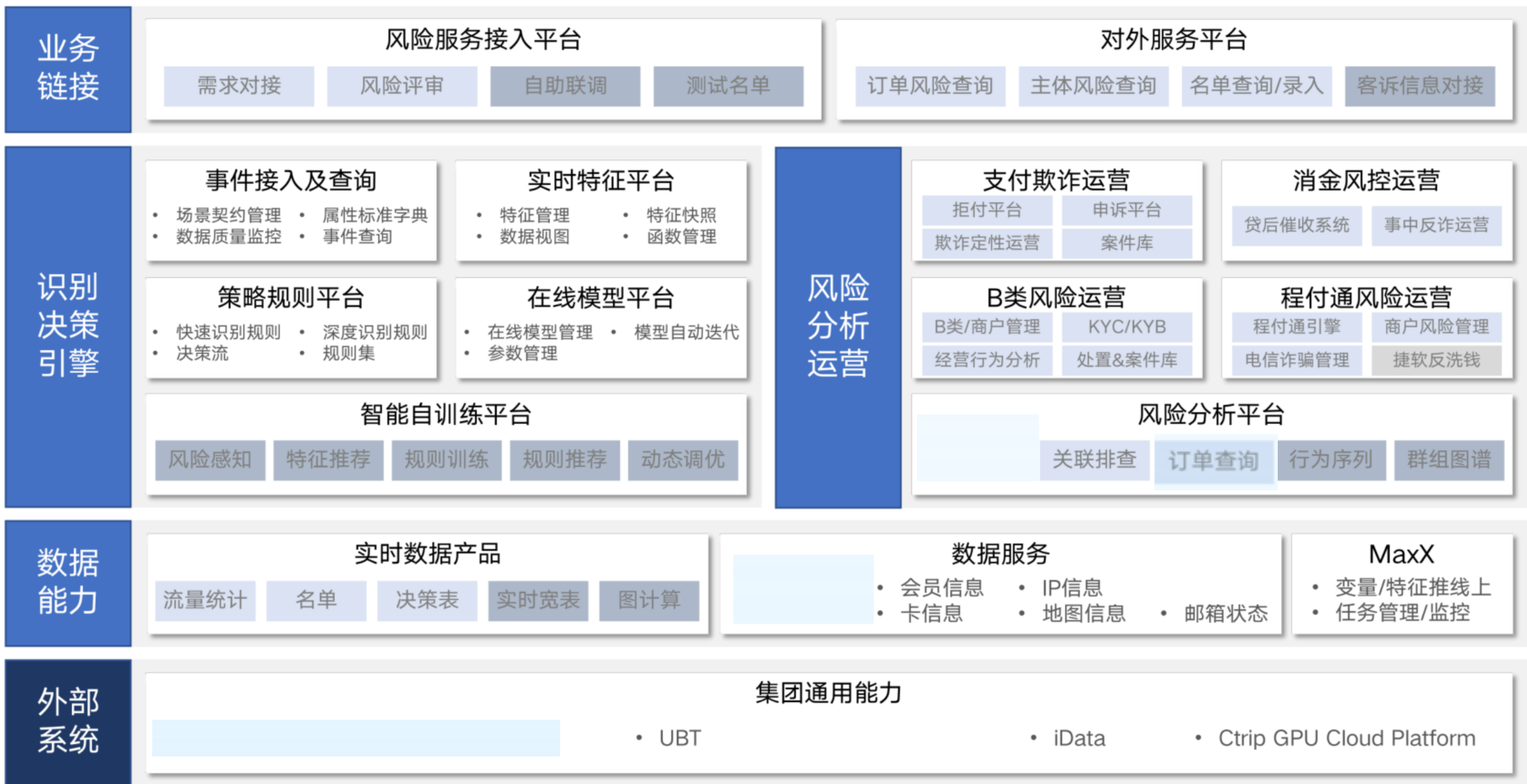
**健壮性强**  
多机房多云DR部署  
对突发流量有较快的响应能力, 自动扩容

**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

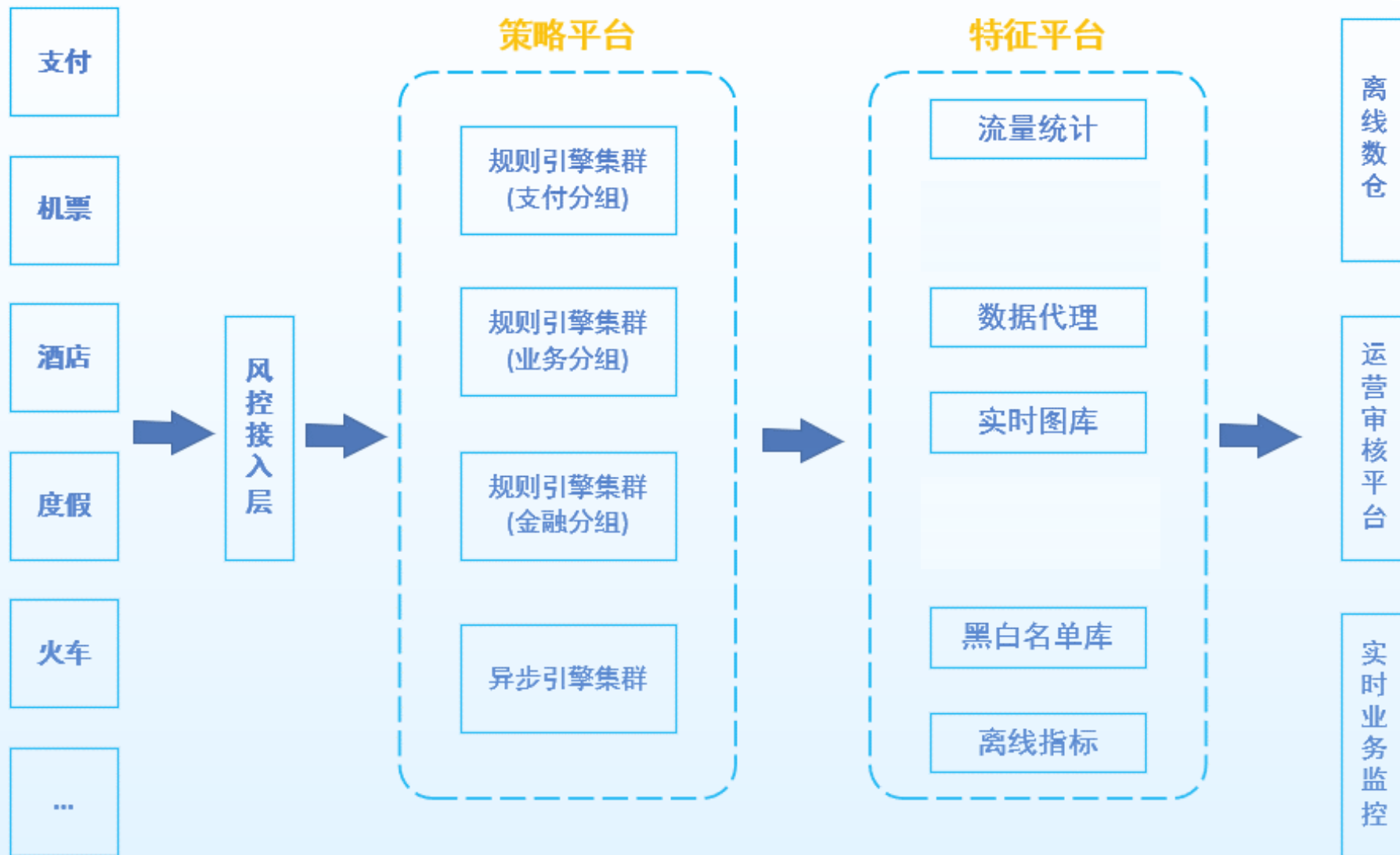




# 风控平台组成



# 风控数据流

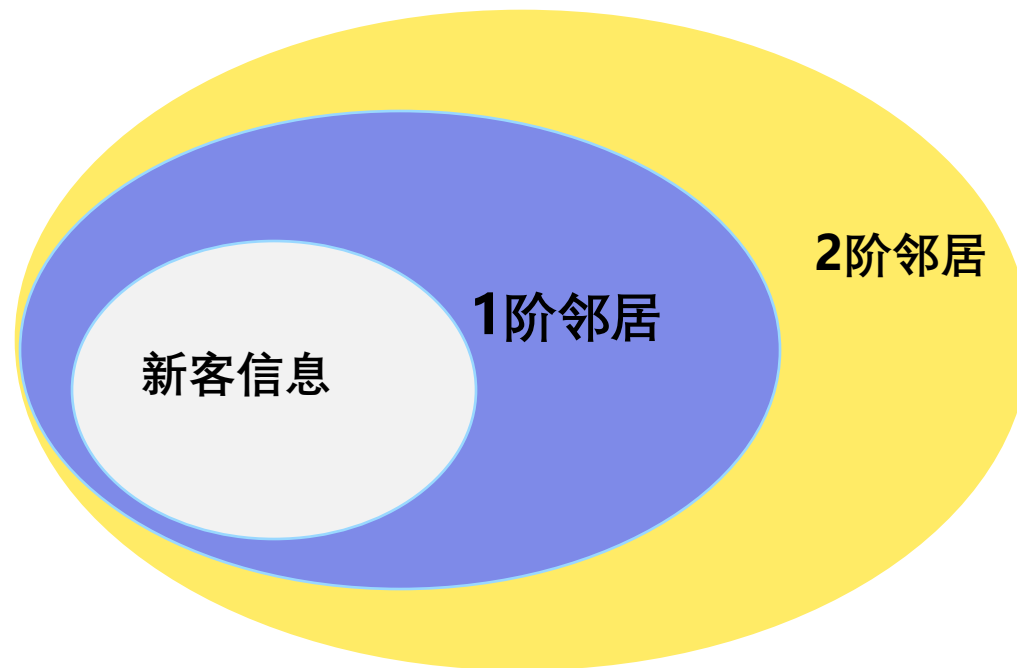


# 2

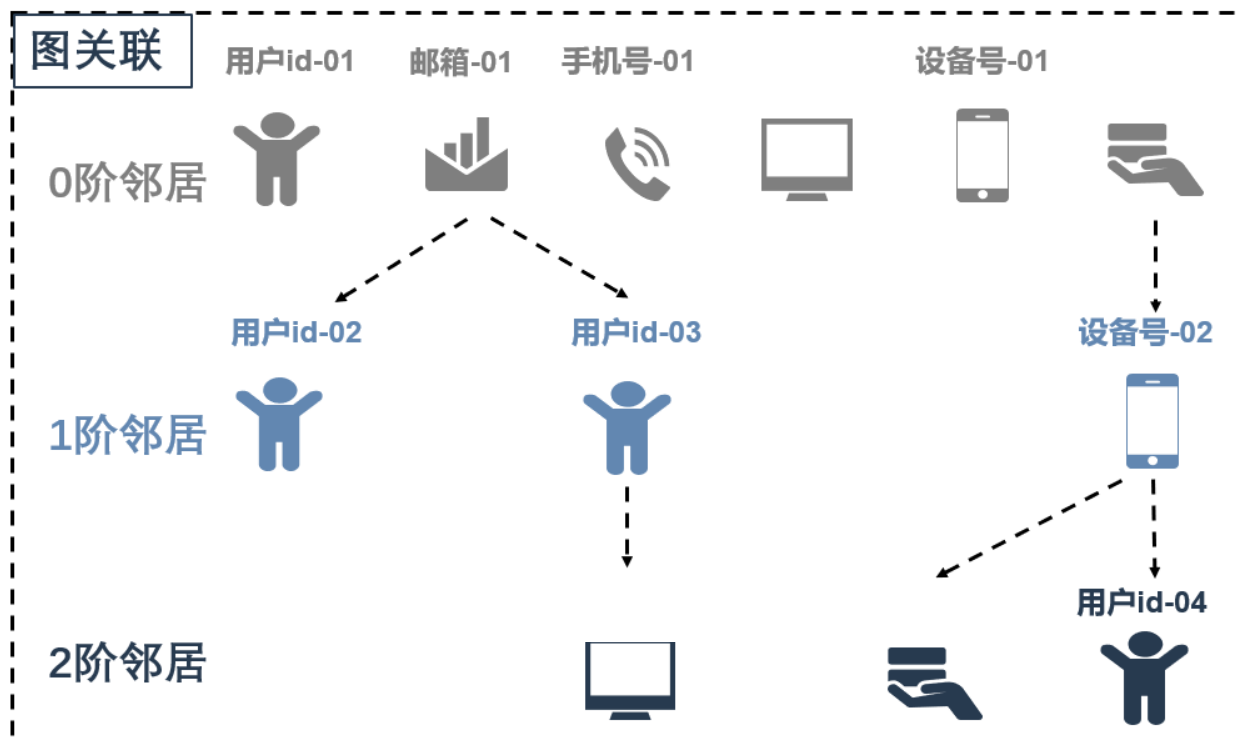
## NebulaGraph在携程风控 的落地实践

## ● 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

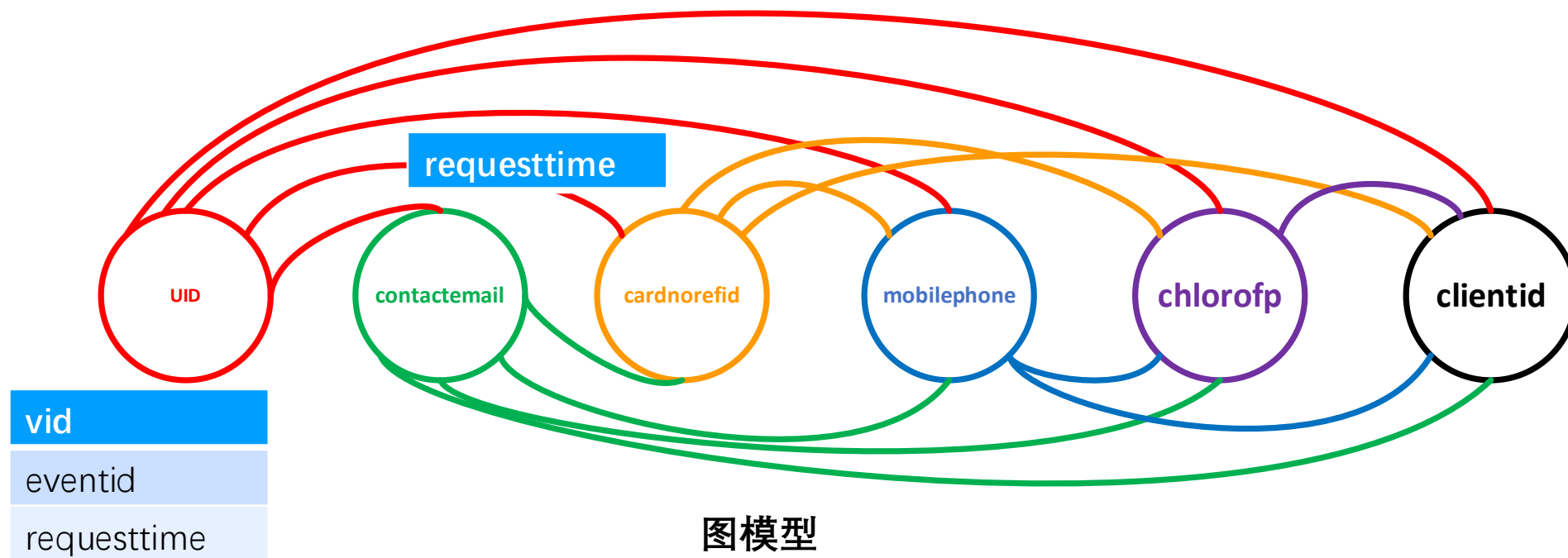


- 对业务行为进行建模, 实时进行图的写入关联
- 支持实时查询当笔交易的**2阶关联**
- 保持**一年内**的实时关联图

● 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



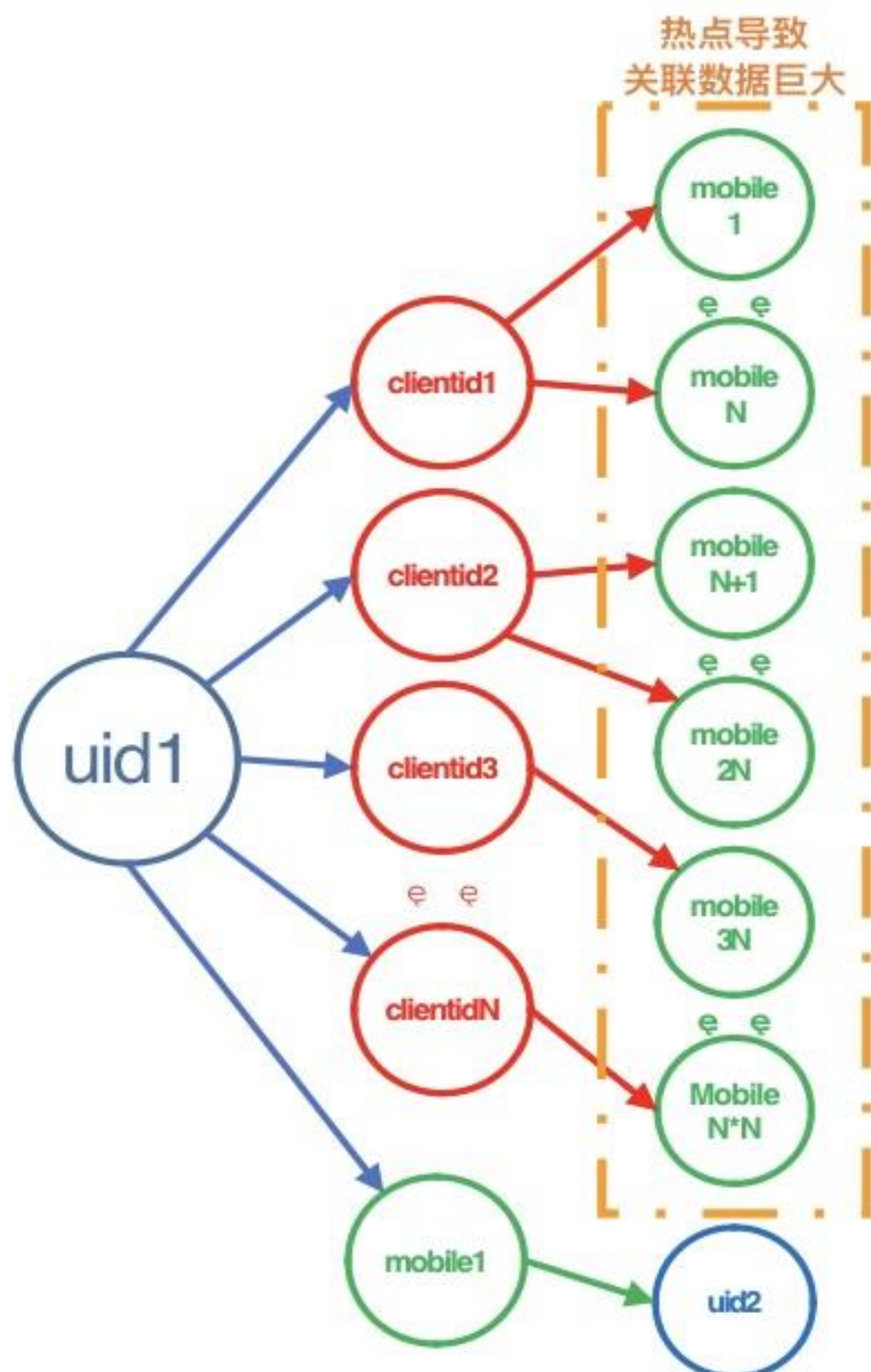
图模型



## 查询逻辑

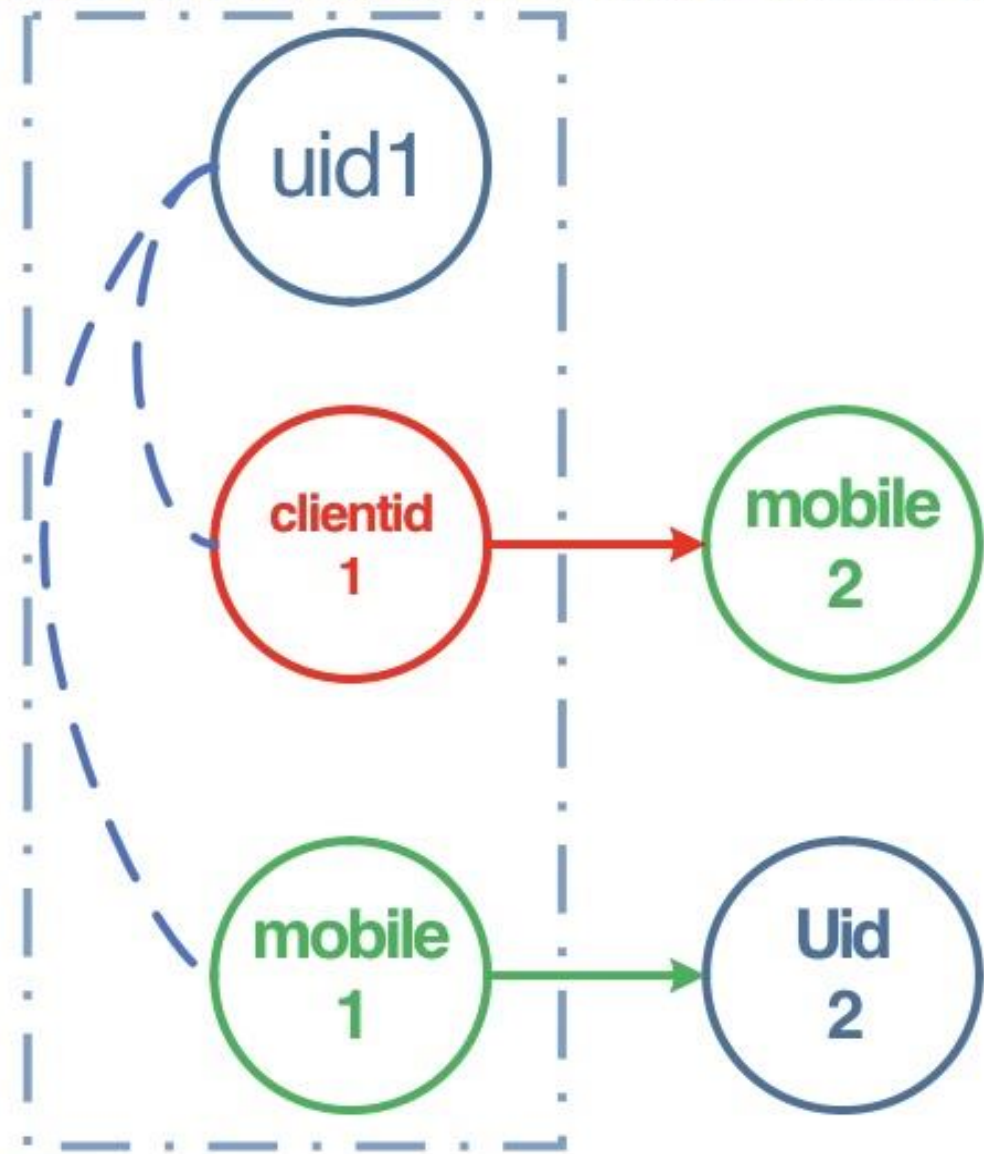
### 难点1

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



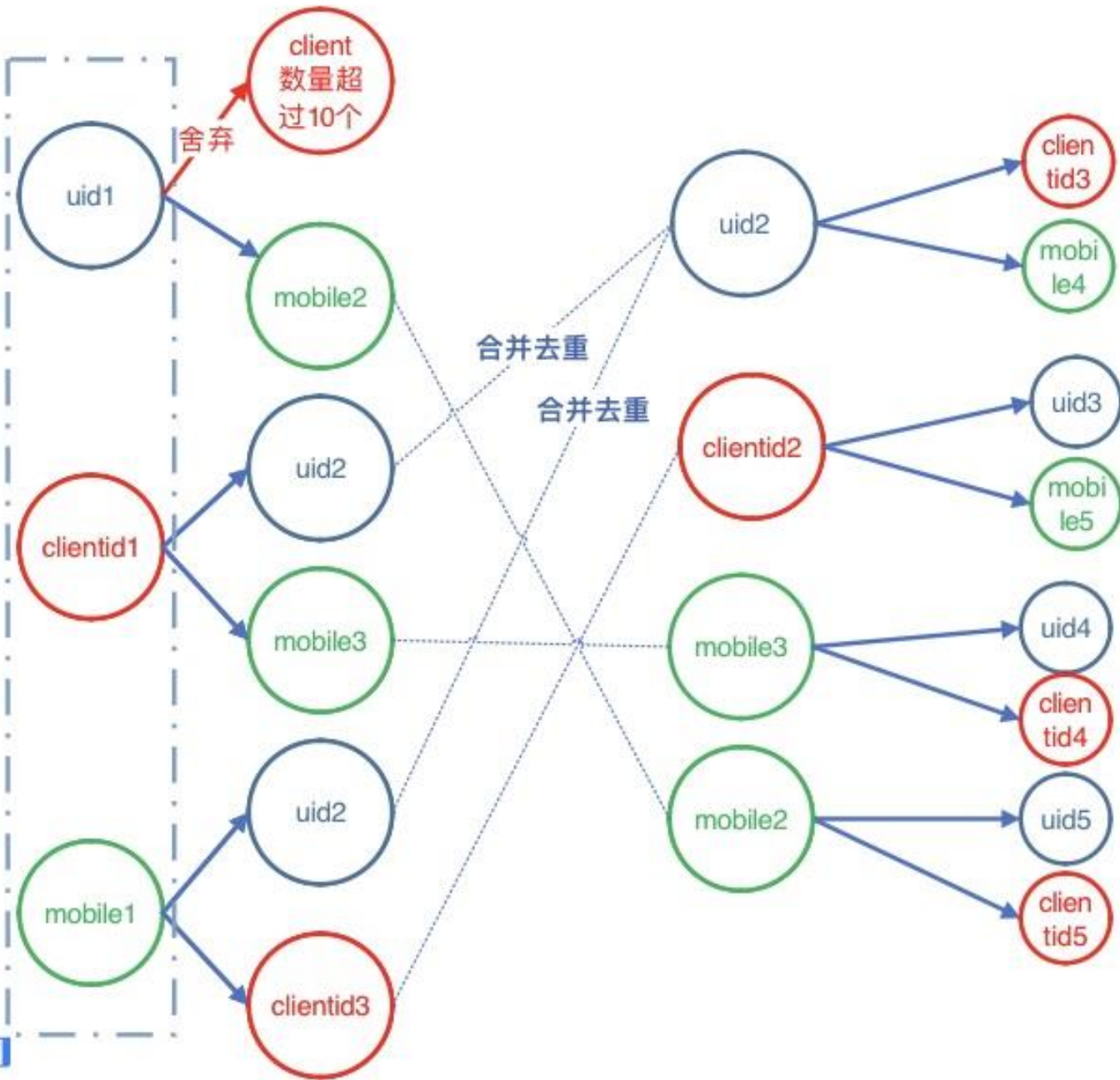
### ● 难点2

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



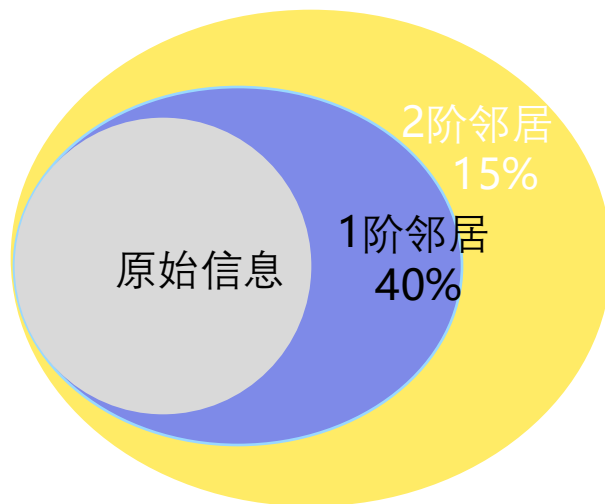




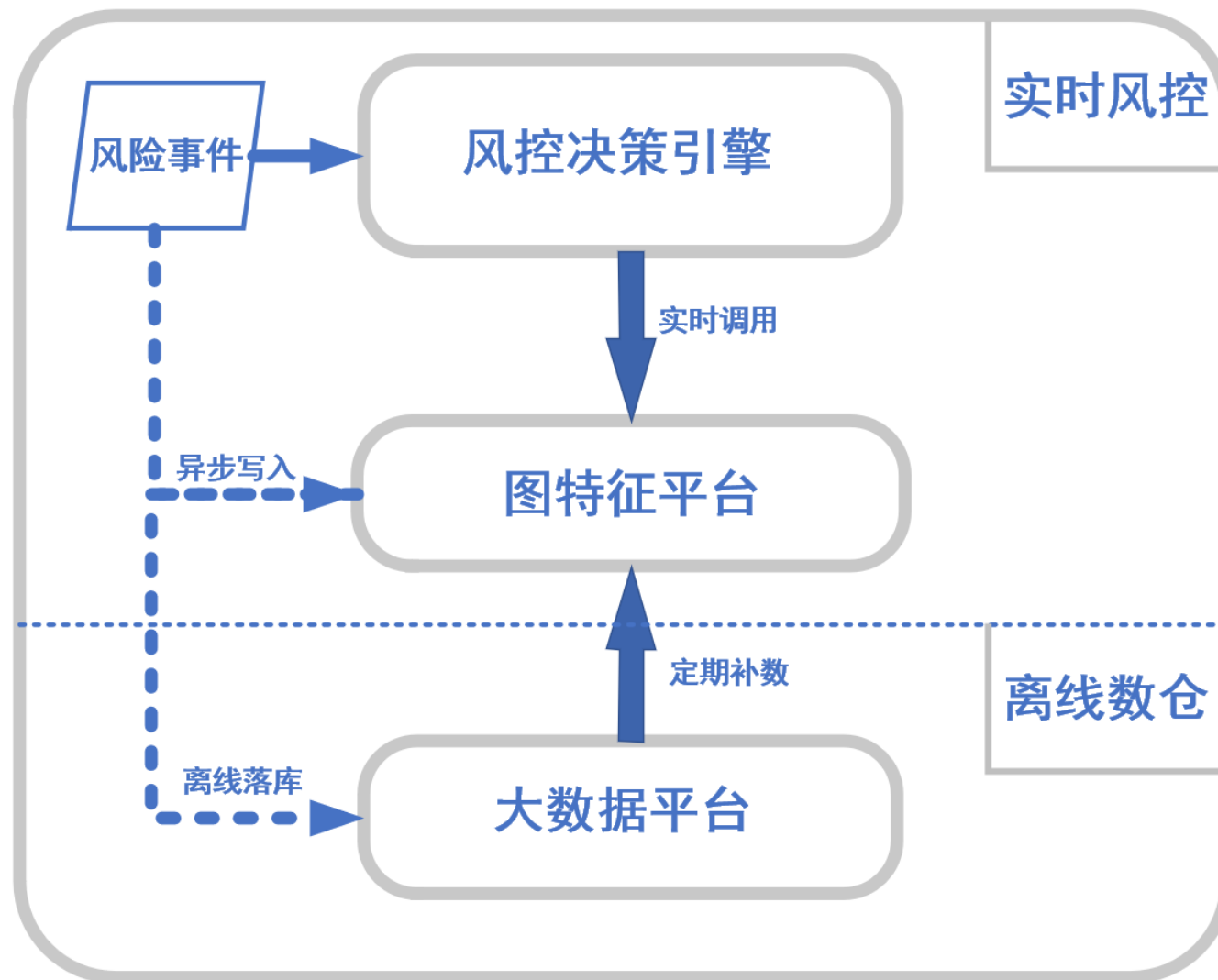
## 项目成果

### 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! QA



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