

金融电子化

# 金融业图数据库建设发展 调研报告

《金融电子化》杂志社有限责任公司

2023年6月

## 版权声明

本报告版权属于《金融电子化》杂志社有限责任公司（以下简称：《金融电子化》杂志社），并受法律保护。转载、摘编或利用其他方式使用本调研报告中的文字、图表、观点，应注明“来源：《金融电子化》杂志社”。未经《金融电子化》杂志社的书面许可，任何单位或个人不得断章取义、增删、曲解本报告内容。违反上述声明者，《金融电子化》杂志社将保留追究其法律责任的权利。

本报告所涉及的数据来源于业内数据采样，采用深度访问、抽样调查相结合的统计方法。由于统计方法本身的局限性，报告数据与市场真实情况可能存在统计误差。《金融电子化》杂志社对其独立研究或与其他机构共同合作的所有研究数据、研究技术方法、研究模型、研究结论及衍生服务产品拥有全部知识产权。

本报告及衍生产品最终解释权归《金融电子化》杂志社所有。

## 前言

随着数字经济的高速发展，以图数据库、图计算、图学习为代表的图技术迎来了蓬勃发展的黄金时期，新技术在赋能数字经济的同时也带来了应用实践上的全新挑战。

作为算力和数据存储的载体，图数据库技术能够实现底层数据的互联互通，充分挖掘数据内在关联，最大化释放数据资产价值，已成为金融业数字化转型的重要基础设施。国家相关部门高度重视，先后出台多项政策，推动图技术在金融行业的应用。中国人民银行《金融科技发展规划（2022—2025 年）》明确指出：运用联合建模、图计算、数据可视化、数字孪生等技术手段，对海量多样化多维度数据资源进行价值挖掘和关联分析，建立面向用户、面向场景的大数据知识图谱和综合分析能力。

因此，金融机构对激活数据要素和数智转型越发关注，图数据库凭借深挖数据价值、创新数据应用、释放数据潜能的关键能力，成为金融机构高效治理、利用数据资产实现业务升级的主要选项。

为更好地了解我国金融业图数据库技术的应用现状及需求，发现行业共性问题，宣传成功经验，我们组织了针对金融业图数据库建设发展的调研工作，从建设背景、建设现状与挑战、趋势与展望等维度，对近百家金融机构进行了深入细致地调研，也获得了各方面的大力支持和积极参与。

通过对调研数据科学的研究与分析，编撰出版了《金融业图数据库建设发展调研报告》，希望可以为金融行业的相关从业者和研究者提供有价值的参考和借鉴，有效推动金融业务创新和发展。

## 编制委员会

主 任 潘润红

副 主 任 杨文艳

成 员 张学峰 李明富 魏斌 郑艺

美术编辑 张芹

《金融电子化》杂志社有限责任公司  
业务策划中心

电 话 (010) 82186848 (010) 82186849

地 址 北京市西城区金融大街 37 号百盛大厦北楼 5 层 539 室

邮 编 100032

## 贡献单位

中国银联

中国农业银行

交通银行

中信银行

华夏银行

广发银行

上海浦东发展银行

浙商银行

建信金融科技有限责任公司

北京银行

秦皇岛银行

阜新银行

辽宁振兴银行

吉林亿联银行

江苏银行

厦门国际银行

青岛银行

郑州银行

汉口银行

湖南银行

广州银行

重庆银行

成都银行

中国工商银行

中国银行

中国邮政储蓄银行

中国光大银行

中国民生银行

平安银行

渤海银行

成方金融科技有限公司

兴业数字金融服务（上海）股份有限公司

河北银行

晋商银行

大连银行

吉林银行

上海银行

杭州银行

齐鲁银行

中原银行

湖北银行

武汉众邦银行

长沙银行

广东华兴银行

四川银行

四川新网银行

贵阳银行

青海银行

河北省农村信用社联合社

长春农村商业银行

浙江农村商业联合银行

江西省农村信用社联合社

青岛农村商业银行

湖北省农村信用社联合社

湖南省农村信用社联合社

广州农村商业银行

深圳农村商业银行

广西壮族自治区农村信用社联合社

四川省农村信用社联合社

贵州省农村信用社联合社

青海省农村信用社联合社

中国人寿保险股份有限公司

中国银河证券股份有限公司

西安银行

北京农村商业银行

辽宁省农村信用社联合社

江苏省农村信用社联合社

宁波鄞州农村商业银行

山东省农村信用社联合社

河南省农村信用社联合社

武汉农村商业银行

广东省农村信用社联合社

广东顺德农村商业银行

东莞农村商业银行

重庆农村商业银行

成都农村商业银行

甘肃省农村信用社联合社

中国人民财产保险股份有限公司

阳光财产保险股份有限公司

浙江创邻科技有限公司

# 目 录

观点摘要	1
第一章 调研背景与方法	4
一、调研背景	4
二、调研内容	4
三、调研对象及方法	4
第二章 图数据库建设洞察	6
一、图数据库知识介绍	6
1. 图数据库概念	6
2. 图数据库技术的发展历史	6
3. 图数据库的优势	8
4. 图数据库的分类	8
二、加快图数据库建设的驱动因素	10
1. 国家政策支持推进建设进程	10
2. 激活数据要素是行业的共性需求	10
3. 图数据库成为解决业务痛点的重要手段	11
三、图数据库市场现状与生态	12
1. 全球市场现状	12
2. 图数据库生态	13

第三章 我国金融业图数据库建设现状	14
一、被调研机构样本统计	14
二、金融业图数据库整体表现	14
1. 图数据库建设发展不均衡	14
2. 图数据库应用场景多元发展	19
三、金融业图数据库选型关注重点	23
1. 性能表现成为金融机构关注焦点	24
2. 安全可控是金融机构的重要考虑因素	27
3. 业务支撑能力是普遍关注点	29
4. 简单易运维受行业青睐	29
5. 用户普遍关注企业在行业的应用能力	32
四、金融业图数据库建设难点	32
1. 人才紧缺、业务场景价值有待验证是建设的主要瓶颈	32
2. 亟待建立与推广图数据库行业标准	36
第四章 金融业图数据库建设与应用趋势	38
一、金融业图数据库建设整体路径	38
1. 自研应用+外采图数据库成为首要建设路径	38
2. 业务发展是驱动金融机构图数据库应用的主要动力	39
二、金融业图数据库建设方向	39
1. 打破数据孤岛，打造企业级图数据库平台	40
2. 图数据库应用场景向多元化、实时化发展	42
3. 图数据建设与应用规模增速迅猛	44
三、金融业期待与展望	46
1. 大规模图数据实时分析需求增加，推动图技术向 Graph3.0 时代升级	46
2. 与 ML、DL 等人工智能技术结合，实现可解释的 AI	48
3. 关注时序数据分析，实现业务动态化分析	49



第五章 图数据库未来发展建议·····	52
一、底层技术自主可控成为关键竞争力·····	52
二、人才是打通图数据库金融行业应用的关键·····	52
三、存储、计算、分析一体化的图平台或成主流·····	53
第六章 金融业图数据库实践案例·····	55
案例 1 农业银行图数据库建设历程·····	55
案例 2 交通银行全栈式知识图谱平台及应用建设·····	59
案例 3 民生银行知识图谱建设与应用·····	62
案例 4 图数据库建设发展思考与实践·····	65
案例 5 基于图数据库的金融知识图谱典型应用·····	67
案例 6 图数据应用建设探索与实践·····	71
案例 7 基于知识图谱的智能风控平台实施案例·····	74
案例 8 图计算在中原银行反欺诈领域的探索应用·····	77
案例 9 阳光财险非车新一代产品工厂·····	81
参考文献·····	83

## 观点摘要

### 数字经济创造发展机遇，国家政策支持图数据库建设

随着通信技术和大数据的快速发展，金融行业所具有的数据资产体量剧增，数据之间的关联关系也变得更加复杂交错。图数据库作为深入挖掘数据信息的有力工具，受到全行业的广泛关注。在国家将高性能计算技术提升到国家战略高度的大背景下，图数据库建设迎来了发展的黄金时代。

2021 年 11 月，工信部发布《“十四五”软件和信息技术服务业发展规划》，在“关键基础软件补短板”专栏中，明确指出“突破全内存高速数据引擎、高可靠数据存储引擎、分布式数据处理与任务调度架构、大规模并行图数据处理等关键技术，推动高性能数据库在金融、电信、能源等重点行业关键业务系统应用。”

### 图数据库在金融行业获得长足发展，但总体建设进度不一

作为创新技术，虽然图数据库技术尚在技术采用周期的早期阶段，但在我国数字化程度领先的金融行业，已获得长足发展。本次被调研的 100 家机构中，79 家金融机构返回了有效问卷，超过半数的机构已在使用图数据库，仅有 2.5% 的机构表示没听说过图数据库技术。

图数据库的使用受到机构规模和区域环境的双重影响，呈现出建设进度不一的特征。其中，国有大型商业银行发挥领跑优势，对图数据库的认知、投入、建设规模均处于行业领先阶段，被调研的 6 家国有大型商业银行中，83.3% 已进入图数据库的使用阶段，10 家股份制银行均已进入图数据库的使用阶段，6 家其他金融机构中，66.7% 已进入使用阶段；32 家城商行中，50% 正在使用图数据库；25 家农信社、农商行中，72.0% 的机构未听说过图数据库，或听说但未使用过图数据库。长三角、珠三角地区的图数据库使用程度最高，超过半数的机构已使用一种及以上图数据库。

## 金融机构图数据库建设规模增速快、技术应用差异化程度高

图数据库能够实现底层数据的互联互通，充分挖掘数据要素的业务价值，已成为金融机构在数字化转型中的关键选择，被越来越多的金融机构所认知并采用。

统计截至 2022 年末金融机构图数据库建设现状发现，已使用图数据库的 41 家机构技术应用差异化程度高：图数据量从 100GB 到超过 10TB，58.5% 的金融机构达 TB 量级，17.1% 超过 10TB；图规模从百万点边到超过百亿点边，主流在 1 亿 -10 亿点边规模（36.6%），14.6% 达 10 亿 -100 亿点边规模，9.8% 的机构建设规模达 100 亿以上；图服务器节点数量 2-3 个占比 39.0%；4-10 个占比 31.7%，12.2% 的机构部署了超过 30 个节点。

统计被调研机构的 2023 年图数据库建设规划发现，图数据库市场增速快：规划中，31.7% 机构的图数据规模达 10TB 以上，同比增长 85.7%；主流依然为 1 亿 -10 亿点边规模（39.0%），24.4% 达 10 亿 -100 亿点边规模，同比增长 66.7%，17.1% 达 100 亿以上点边规模，同比增长 75.0%；图服务器节点数量 2-3 个占比 22.0%，同比降低 50.0%，4-10 个占比 36.6%，同比增长 15.4%，14.6% 的机构计划部署超过 30 个节点，同比增长 20.0%。这是由于一方面金融机构业务量的迅速增长导致数据量激增，对大数据的关联存储查询性能提出了更高需求；另一方面，金融机构的数字化转型压力在不断增加，发挥数据资产价值的需求更加强烈，促进了以图技术为驱动的业务应用创新。

## 大数据查询性能与自主可控是金融机构选型的首要关注点

随着数据规模急剧增长，存储处理的数据类型、数据维度多样化、复杂化，图数据库技术有助于在即时查询，并发操作，关联关系等方面大幅赋能业务数据潜在价值的释放。金融业对于图数据库产品的技术价值关注度主要聚焦于存储、查询、分析能力上，87.8% 的被调研机构将图数据库的大数据处理性能看作最重要的考量指标。

此外，随着政治环境的不断变化、金融机构对安全、自主可控提出了更高的要求。48% 的被调研机构将产品闭源、代码安全性保障、能够规避知识产权风险或是否“100% 自主可控”列为重要考量因素。50% 以上被调研机构均认为“100% 自主可控”是重要考量指标，其中 40% 国有大型商业银行将其作为首要关注的指标。60% 的被调研机构明确表示会选择自主可控的图数据库，其中所有被调研的股份制银行均表示必须为自主可控的图数据库。

## 金融机构图场景建设方向明确，企业级图平台为首要建设目标

金融机构在规划未来建设方向时，对于整合全行数据资源，打通各部门“数据孤岛”，实现统一的数据管理、关联挖掘、业务洞察表现出了极高的迫切性。

在图技术场景建设方面，被调研机构更聚焦传统数据库不能有效支持的创新应用，超 90% 的被调研机构规划建设企业关系图谱库、实时交易反欺诈应用、反洗钱应用。超过 60% 的被调研机构有零售营销图谱及数据血缘场景的建设计划。虽然当前图技术主要由创新应用场景引入，但金融机构也明确表示希望形成跨单一场景的平台能力。调研结果显示，在有计划部署图数据库的 49 家金融机构中，企业级图数据库平台成为首要规划建设的场景，80% 的机构表示有相关建设计划。

## 商业图数据库优势明显，单一供应商成为主流选择

图数据库及其应用市场正处于快速发展迭代的过程中，市场内相关厂商较多，商业图数据库产品在性能表现、核心代码安全、综合运维管理成本，以及成熟的服务团队等方面都具有显著优势。相对于从 0 到 1 进行图数据库自研，或是基于开源图数据库自行搭建系统，金融机构更倾向于底层图数据库外采、上层应用自研的建设思路。

在 49 家计划部署图数据库的金融机构中，仅有 5 家机构表示会以自研为主，近 90% 的机构表示会选择与商业图数据库厂商合作，其中 73.0% 表示会选择单一商业图数据库供应商，以降低管理与运维复杂度。

## 打通应用“最后一公里”，成为图数据库技术发展的关键点

在计划部署图数据库的 49 家被调研机构中，图数据库厂商的实际落地案例成为金融机构的首要考量因素，95.9% 的被调研机构均表示更倾向于与已有金融机构合作案例的图数据库厂商开展合作，股份制银行则尤其看重大型银行的合作案例。此外，83.6% 的被调研机构表示会关注厂商是否具备完善的实施和售后支持。

图技术专业人才的稀缺已经成为制约金融行业图数据库应用的重要因素，78.5% 的金融机构表示缺乏相关的技术专业人才。国有大型商业银行作为图数据库的深度用户，对业务、技术，以及技术业务结合的复合型人才均表现出了极高的需求。因此，被调研机构均表示期待有更多机会参与图数据库相关活动以获得技术指导，尤其对业务场景的具体解决方案的支持交流有着强烈需求，希望通过技术交流打通图技术应用“最后一公里”。

## 第一章 调研背景与方法

### 一、调研背景

当前，金融业积累的数据呈指数级增长，新的数据源层出不穷，且数据丰富程度增加，数据间的关联性增多，传统的对小量、单维度、静态化数据的分析已无法满足金融业数字化转型发展的需求。传统的关系型数据库无法满足高效、可扩展、复杂数据管理和分析的需求，存在数据处理效率低、分析应用效果弱的问题。

图技术以“点”“边”为数据结构直观简洁地描述业务实体之间的复杂关联关系，以复杂关联关系的存储、查询、计算性能优化为第一设计原则，能够高效地挖掘多源异构数据间潜在的内在关联、最大化释放数据资产价值，提升金融机构对客户的全局洞察力，激发金融机构的业务创新能力。

面对挑战和新的发展机遇，金融机构逐步落地图数据库应用，打造数据互联网络效应，实现金融科技创新，赋能业务发展。为深入了解我国金融机构图数据库技术应用现状及发展趋势，开展本次调研。

### 二、调研内容

本次调研工作主要围绕金融机构在图数据库领域的布局、应用规模、典型业务场景，以及在实际运用遇到的问题和未来发展规划等信息进行采集分析，力求客观反映当前国内金融机构图数据库的建设、应用和发展情况。

### 三、调研对象及方法

参与调研的金融机构包括：国有大型商业银行（简称国有大行）、全国性股份制商业银行（简称股份制银行）、城市商业银行（简称城商行）、农信社（农商行）及其他金融机构（银联、证

券和保险机构)。调研采用定量与定性结合的方法。定量方面,通过向 100 家金融机构发出份问卷,收到有效问卷 79 份,涵盖了长三角、珠三角、京津冀、中部、西部、东北等六大区域的金融机构。定性方面,与 13 家金融机构专家进行交流,详细调研了各家图数据库的应用现状。

## 第二章 图数据库建设洞察

### 一、图数据库知识介绍

“图”是以图论（Graph Theory）为理论基础、以点（实体）和边（关系）为核心逻辑、描述现实世界中个体和个体之间复杂关系的数据结构，对现实世界的内在关系具有简洁、直观的表达能力。

图技术泛指一切研究事物和事物之间的关系，描述、刻画、分析和计算事物之间关系的技术，用于从图结构数据中挖掘出有价值的知识或规律来指导业务决策，如风险评估、事件溯源、因果推理和影响分析等。图技术可大致分为三大类：一是处理图数据存储查询的图数据库技术；二是对图数据进行分析 and 计算的图计算技术；三是对图数据进行可视化展示的图可视化技术。实际应用中，三者的能力呈现融合趋势。

图数据库是以点、边为基础存储单元、以高效存储、查询图结构数据为第一设计原理的数据管理系统。图数据库非常善于处理海量关联数据，对于揭示数据之间的内在联系具有很大的性能优势。

#### 1. 图数据库概念

图的概念对于图数据库的理解至关重要。图（Graph）源于数学中的图论，是一组顶点和边的集合，“顶点”表示实体，“边”表示实体间的关系。图描述的是现实世界中一组对象以及这些对象之间存在的一组关联关系。

图数据库是使用图结构存储数据、实现图语义查询的数据库管理系统，它通过“顶点”“边”和它们的“属性”来表示并存储数据。不同于传统数据库，数据间的关系以表与表之间的外键的形式隐形、间接存在，针对关系的查询必须在运行时进行具体化操作，在图数据库中，数据间的关系和数据本身同样重要，它们被作为数据的一部分显性存储，关系上可添加方向以及属性对其进行进一步描述。这样的存储设计使图数据库能够直观、快速地响应复杂关联关系以及关系特征的查询，是存储、查询、分析高度互联数据的最佳技术方案。

#### 2. 图数据库技术的发展历史

以属性图为核心数据模型的现代图数据库从诞生到大规模应用大致分为三个发展阶段（见图1）。





▲ 图 1 图数据库技术发展的三个阶段

数据来源：《金融电子化》杂志社，2023 年 3 月

Graph 1.0: 单机原生图数据库

2002-2010 年，以 Neo4j 为代表的支持事务性的原生图数据库的发布，标志着图数据库技术的诞生。原生图数据库采用“免索引邻接”的设计，让图遍历的性能仅与被查询的顶点关联到的局部数据量相关，而不会像传统数据库那样随全局数据体量的增加而降低，实现了关联查询效率的极大提升。这个阶段的图数据库扩展性差，受限于单机的存储计算资源瓶颈，应用主要集中在知识图谱、股权图谱、数据血缘等小型静态数据的分析场景。

Graph 2.0: 分布式非原生图数据库

2010-2015 年，随着大数据时代的到来，企业收集的数据的丰富程度与体量急速增长。图数据库应用开始从静态参考数据扩展到基于行为的核心交易数据，如资金流向、信贷申请、消费及生产关系上。此时，系统的横向扩展性成为行业刚需，是否支持大规模数据的分布式存储成为人们关注的重点。以 JanusGraph、ArangoDB 等为代表的分布式图数据库不断涌现，这些数据库底层基于分布式的开源 NoSQL 存储，实现了图数据的分布式存储和加工，一定程度上解决了单机图数据库在大数据量下的性能和扩展性问题。由于底层基于分布式的 NoSQL 存储，Graph 2.0 时代的图数据库产品的横向扩展性有长足提升。但由于底层的键值数据库、列式数据库等在内核设计上并不以“关系”的表达和处理作为优化重点，图遍历的查询性能较 Graph1.0 时代的原生图数据库有较大差距，无法满足大数据实时更新、查询、计算的需求，这个时代的图数据库技术多应用于大数据的离线分析场景。

Graph 3.0: 原生分布式图数据库

2016 年开始，随着市场认知度的加强，图数据库技术的技术成熟度不断发展，应用场景也



不断被拓宽、升级到了基于行为、事件、物联网大数据的实时处理、分析场景。为满足大数据量级下的查询返回效率，在快速变化的商业环境下提供实时的商业智能，诞生了以 Galaxybase、Tigergraph 为代表的原生分布式图数据库，有效解决了大图数据深链查询效率和系统横向扩展性的双重需求。

同 Graph 2.0 时代的产品相比，第三代图数据库不再依赖其他 NoSQL 分布式存储系统，直接控制了底层的数据存储与分布式数据通讯机制，实现了分布式的原生图存储，因而大幅提升了大图数据的查询、处理性能。此外，由于图遍历的随机游走特性，大图数据的复杂查询任务往往涉及对不可以预测的数据量的分布式处理，查询优化需要复杂算法支撑，Graph 3.0 时代的图数据库同时增加了分布式并行计算的能力，以支撑实时图分析需求。

大规模原生图存储、分布式并行计算能力正逐渐成为图数据库行业的主流。

### 3. 图数据库的优势

相比传统的关系型数据库，图数据库有诸多优势：

#### 更直观、灵活的模型

图数据模型直接还原业务场景，相比传统数据模型更直观、灵活，能够很好应对数据关系动态变化的场景，提升产品与研发的沟通效率，降低数据模型运营、变更成本。

#### 更简洁的查询语言

图数据库的查询语言在关联查询中更简洁，以通用的 Cypher 图查询语言为例，复杂关联查询的代码量比 SQL 大幅降低，提升应用系统开发效率。

#### 更高效的关联查询

图数据库在处理关联性强的数据及天然的图问题场景时具有强大的关联查询性能优势。传统关系型数据库在进行关联查询时需要做昂贵的表连接（Table JOIN）操作，涉及大量的 I/O 操作及内存消耗，性能存在很大瓶颈。而图数据库专门针对关联查询进行优化，通过数据存储模式的重构，防止局部数据的查询引发的全局数据遍历，可以数万倍提升关联查询的效率。

### 4. 图数据库的分类

图结构数据由具备多种属性的“点”“边”构成，且“点”“边”间存在关联复杂，图结构数据往往呈现出复杂性、多样性和动态性的特征。图数据库的核心操作单元是图遍历，即通过不同的“边”关系查找给定“点”的多跳、不同属性的邻居点。这是所有图查询与图计算的原子操作，它的效率决定了图数据库的整体性能。

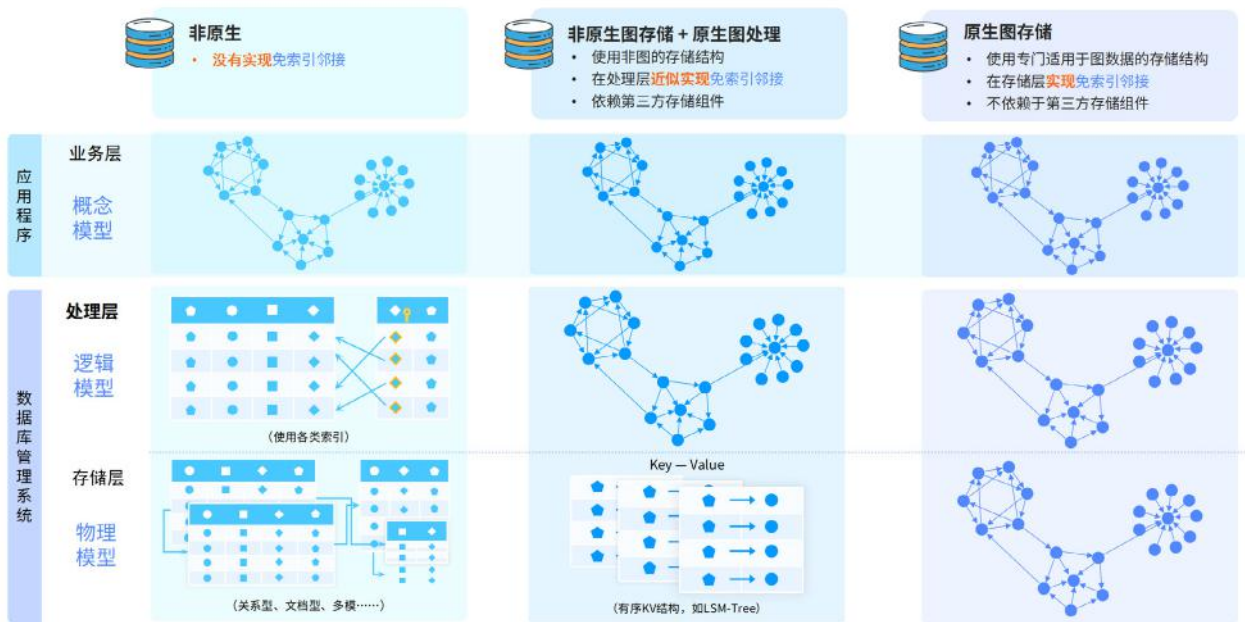
“免索引邻接” (Index-free Adjacency) 是一种数据组织形式，它能保证一个点和与它直接相邻的边被存储在一起，无需依赖其他索引类的数据结构即可通过图中的任意给定点直接访问它所

有的相邻边，让找到相邻点的时间成为一个不依赖整体数据量的常量开销。可以说图数据库的技术关键便是在于通过“免索引邻接”实现图查询效率的指数提升。

根据实现免索引邻接的程度及方式不同，相应的图数据库的查询速度和数据的读写性能也会出现巨大的差距。因此，依据不同图数据库在内核设计上是否实现了免索引邻接，以及在什么层面上实现的免索引邻接，可以将图数据库分为三类（见图 2）。

完全非原生。这种结构未实现免索引邻接，它在数据层使用关系型数据库、键值数据库、文档或者其他多模数据库作为存储，在处理层通过多表链接及字段索引实现关联关系的查询，仅在业务层以图的方式进行呈现。该存储架构面对数据量小、关联跳数少、表关系简单、静态的场景也能良好支撑，但在数据量大、关系复杂的场景下查询效率会急速下降，还会带来大量表的大量全局索引的维护工作，对数据关系动态变化的场景运维成本很高，正在被主流图数据库市场淘汰。（见图 2）

非原生存储。该内核设计方式在数据层使用 Key-value 等非原生的存储结构，在处理层近似实现免索引邻接。由于存储层并没有完整的图语义支持，这种架构会导致图查询和计算时，存在存储层与处理层之间从非图模型到图模型转换的额外开销，产生性能损耗。在查询的点边数量较小，关联关系链路不深的场景中，非原生数据库性能尚可，但是当点边数量大，查询链路深，则需要



▲ 图 2 图数据库的分类

数据来源：《金融电子化》杂志社，2023 年 3 月

面对性能下降。此外,在处理层近似实现免索引邻接的过程受限于底层第三方存储系统的实现机制,还会导致查询性能稳定性、数据一致性差等问题。(见图 2)

原生图存储。原生图存储不依赖第三方关系型或 NoSQL 的存储系统,直接在存储层实现免索引邻接,实现复杂图数据遍历的极致性能。在原生图数据库上实现的图处理层,保持了存储与处理层的数据模型的一致性,降低了不同数据模型之间转换的处理成本,能够更好地实现存储与运算系统的无缝联动,降低非原生架构下不同系统间的黑盒沟通成本,最优化图处理性能。(见图 2)

## 二、加快图数据库建设的驱动因素

### 1. 国家政策支持推进建设进程

当前,高性能图计算技术已经上升到国家战略高度。国家陆续出台多重政策,从技术前沿性和市场价值出发,指引图数据库的加速发展。

近年来,国家大力支持信息技术应用创新产业,鼓励图数据库处理技术研发和产品应用。2021 年 11 月,工信部发布的《“十四五”软件和信息技术服务业发展规划》,在“关键基础软件补短板”专栏中明确指出需要:“突破全内存高速数据引擎、高可靠数据存储引擎、分布式数据处理与任务调度架构、大规模并行图数据处理等关键技术。推动高性能数据库在金融、电信、能源等重点行业关键业务系统应用。”相关政策强有力地激发国内企业对于图数据库这项新兴技术的投资和研发,并不断激发各个行业、领域开展对图数据库技术的应用探索。

### 2. 激活数据要素是行业的共性需求

习近平总书记指出:“充分发挥海量数据和丰富应用场景优势,促进数字技术和实体经济深度融合,赋能传统产业转型升级,催生新产业新业态新模式,不断做强做优做大我国数字经济。”

数据作为新型生产要素,是数字经济时代区别农业、工业时代的关键特征。自党的十九届四中全会首次将数据列为生产要素以来,政府陆续出台《“十四五”数字经济发展规划》《关于加快建设全国统一大市场的意见》《要素市场化配置综合改革试点总体方案》、“数据二十条”等一系列顶层设计和落地举措,着力破解数据要素有序流通的体制机制障碍。

中国人民银行《金融科技发展规划(2022-2025 年)》中强调,充分释放数据要素潜能,需要深刻认识数据要素重要价值,制定企业级数据规划和发展战略,深化数据综合应用。并且明确提出“运用联合建模、图计算、数据可视化、数字孪生等技术手段,对海量多样化多维度数据资源进行价值挖掘和关联分析,建立面向用户、面向场景的大数据知识图谱和综合分析能力,发挥数据和技术双轮驱动作用”。

随着信息时代的快速发展，金融机构内部完成了海量数据的沉淀，大规模数字化的客户互动和业务进程的背后会产生海量数据，要求企业对于海量数据的管理、关联、挖掘进行更加深入的探索，实现经营策略从“经验判断”向“数据说话”，组织发展从“产能驱动”向“数据驱动”的转变。

图数据库作为数据时代的重要基础软件之一，能够通过数据关联关系挖掘，全面连接数据要素，激活数据资产价值，赋能金融机构数据应用从“能用、会用”向“好用、爱用”转变，打造数据驱动的全局运营决策能力，帮助金融机构实现数据资源的资产化、价值化。

### 3. 图数据库成为解决业务痛点的重要手段

随着新一代互联网技术的发展，金融机构朝着更全面、深入、便捷的智能化方向发展，内部业务模式日趋线上化、平台化、实时化、场景化。而数字化运营背后，风险防控、智能营销的全新挑战随之而来。

金融风控领域，一方面，金融机构内部传统风控手段出现效率低、精度差、成本高等问题，难以应对互联网技术下风险加速传播和交叉感染的难题；另一方面，犯罪分子在信息技术的加持下日趋职业化、集团化，作案手段多样化，难以直观定位。因此当前金融风险呈现出规模化、隐蔽性、快速性、动态变化的特征。

而在营销场景，金融行业随互联网经济获得的流量红利逐渐见顶，竞争加剧，正逐步转向对已有客户深度运营阶段。传统客户管理系统中客户渠道多、维度少，产品同质化严重，很难做到客户的精准营销，实现差异化运营，客户黏性低、流失率高；并且金融机构线上线下对客渠道割裂，导致内部数据打通难，时间、人力、经济成本较高。

因此，金融机构亟需抓住数字化、网络化、智能化融合发展的机遇，全面推进数字化转型，推动数字技术从“支撑使能”向“价值赋能”变革，为数字经济发展提供强劲的金融支持。中国人民银行《金融科技发展规划（2022-2025年）》中将激活数字化运营新动能定为关键任务，提出“构建集成数据整合、提纯加工、建模分析、质量管控、可视化交互等功能的综合型数据中台，打造科技赋能，数据驱动、业务联动的企业级数据服务能力中枢”。

图数据库，一种基于图论的新型数据管理及分析计算系统，擅长对海量、复杂、多变的数据进行深度、实时分析和计算，已成为金融机构在互联网时代的业务创新发展的核心引擎。基于图数据库，金融机构能够高效整合交易、客户、风险、运营等数据以及第三方数据，打造立体化的多维 KYC 完整视图，构建全行一张图，实现全客户、全渠道、全流程、全场景的数据智能融通，增强客户跨业务、跨产品、跨时间的风险识别及动机识别能力，并针对不同场景的不同客户定制化营销策略，助力客户全生命周期的精准营销，真正做到打通业务条线壁垒、解构业务逻辑、沉



淀通用业务能力，赋能金融机构业务能力实现质的飞跃。

### 三、图数据库市场现状与生态

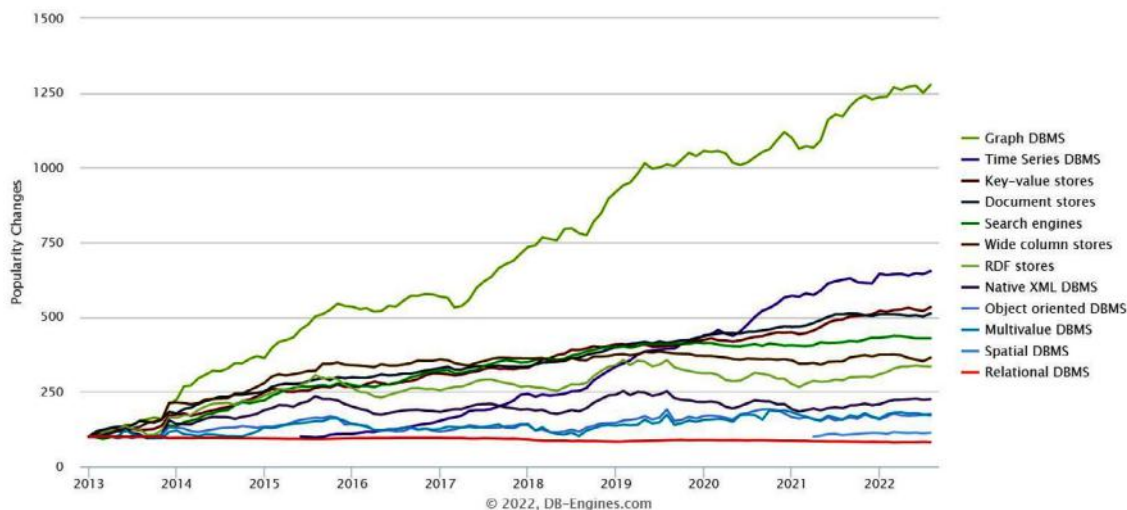
#### 1. 全球市场现状

当数字经济、产业数字化转型进入深化发展阶段时，数据作为产业发展的核心资产，被赋予新的角色和使命。企业更为关注数据资产的利用，图技术在处理复杂关联数据上的性能优势决定其在未来数据资产价值变现的场景有广阔的市场空间。

Gartner 在《2022 年图数据库管理系统市场指南》中预估，“到 2025 年，包括图数据库管理系统（DBMSs）在内的图技术市场将增长到 32 亿美元，年复合增长率为 28.1%”。并预测，“到 2025 年，图技术将用于 80% 的数据和分析创新，高于 2021 年的 10%，促进企业快速决策”。

在数据体量日益膨胀和数据间关联关系日益复杂的今天，越来越多的企业和研究机构都开始关注图数据库的应用。数据库门户网站 DB-Engines 数据显示，2013—2022 年图数据库的热度增长远远高于其他类型数据库（见图 3）。

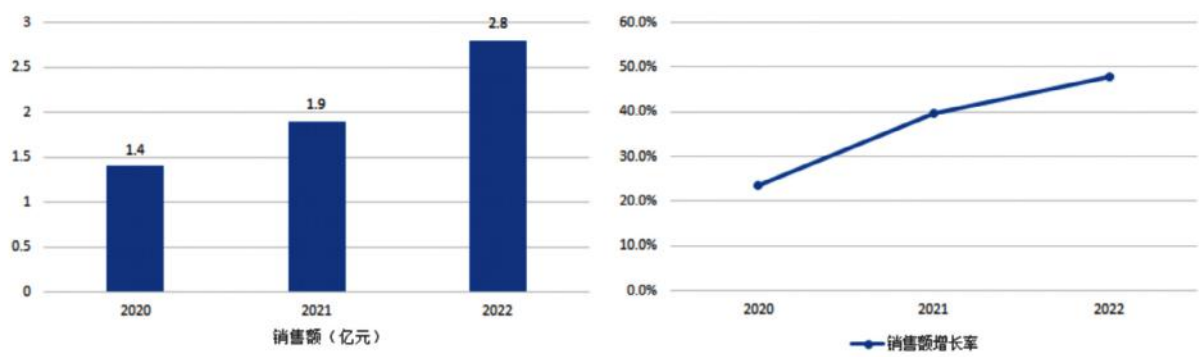
Complete trend, starting with January 2013



▲ 图 3 各类数据库近十年来热度变化

数据来源：DB-Engines.com，2022 年 12 月

当前，基于图数据库技术日益增长的普及与使用，各国竞相布局图数据库赛道，全球图数据库市场已形成初步规模。中国的图数据库市场在政策和市场需求的共同推动下，正处于快速发展的起步阶段，市场中充满机遇和竞争（见图 4）。



▲ 图 4 2020-2022 年中国图数据库市场规模及增长

数据来源：赛迪顾问，2023 年 4 月

2. 图数据库生态

2019 年末，国内图数据库市场加速发展，各类图数据库厂商在市场中亮相，图数据库迅速在各领域展开应用。

现阶段商业图数据库厂商大致可分为三类：一是公有云厂商基于自身业务需求及技术架构开发的数据库产品，并逐步由内部能力提供向外部赋能，典型代表厂商为 AWS、腾讯云、阿里云、百度等，它们占据了图数据库公有云的主流市场；二是新兴的商业图数据库创业公司，其图数据库产品最大特点是高性能与更贴近客户需求的服务能力，典型代表厂商为创邻科技、TigerGraph 等；三是传统数据库厂商、应用开发商关注到图数据库的发展趋势，为了弥补自身产品矩阵在图数据库方面的短板而打造的产品。这类厂商分布在图技术产业链的不同上下游位置，共同推动图技术的发展与应用（见图 5）。



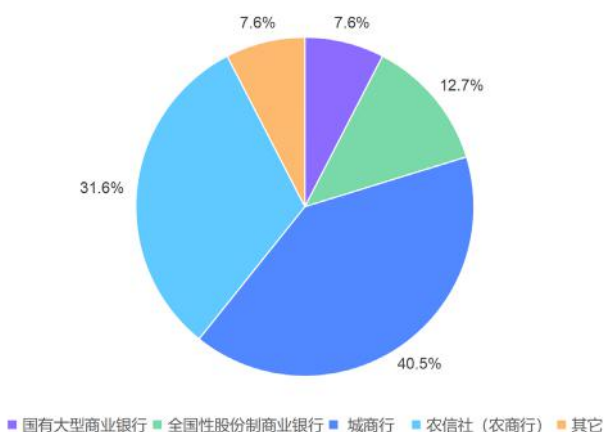
▲ 图 5 图技术产业生态

数据来源：CB Insights 中国，2022 年 12 月

## 第三章 我国金融业图数据库建设现状

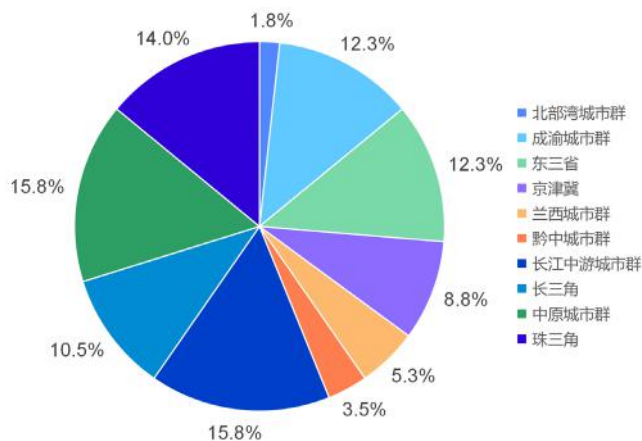
### 一、被调研机构样本统计

本次调研，我们共发出 100 份问卷，收集有效问卷 79 份。从机构类型来看，包括 6 家国有大型商业银行、10 家全国性股份制商业银行、32 家城市商业银行、25 家农信社（农商行）以及 6 家其他金融机构。从分布地域来看，57 家城商行、农信社（农商行）按地域划分，1 家银行来自北部湾城市群、7 家来自成渝城市群、7 家来自东三省、5 家来自京津冀、3 家来自兰西城市群、2 家来自黔中城市群、9 家来自长江中游城市群、6 家来自长三角、9 家来自中原城市群、8 家来自珠三角。（见图 6-1、6-2）。



▲ 图 6-1 调研对象机构类型样本描述

数据来源：《金融电子化》杂志社，2023 年 3 月



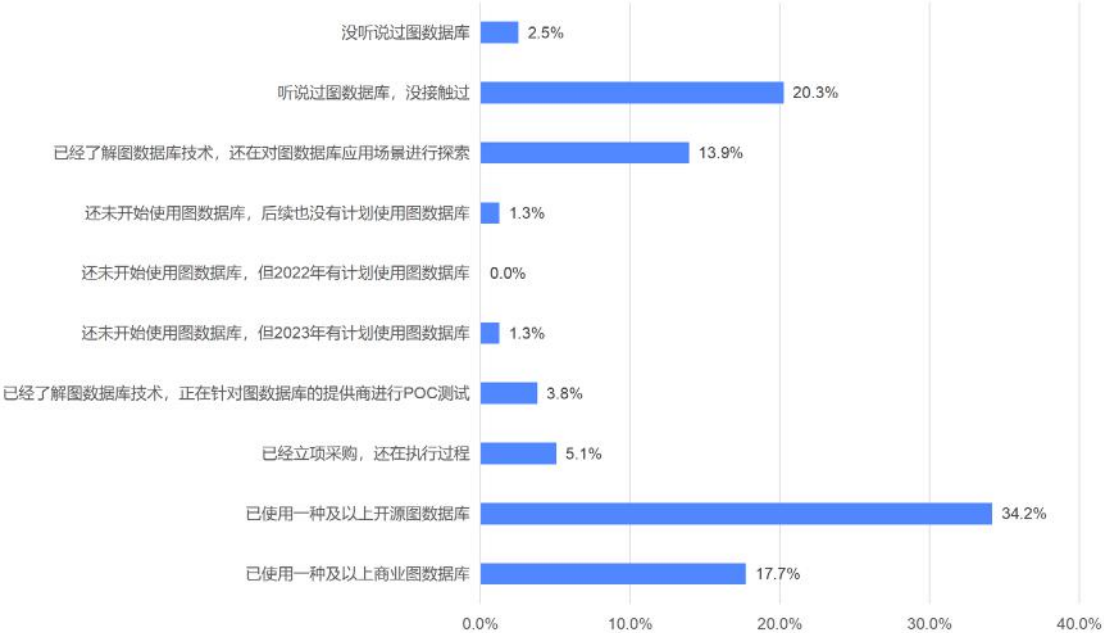
▲ 图 6-2 调研对象分布地域样本描述

数据来源：《金融电子化》杂志社，2023 年 3 月

### 二、金融业图数据库整体表现

#### 1. 图数据库建设发展不均衡

调研数据显示，图数据库技术在金融行业已经开始进入大众市场的视野。被调研的金融机构中，仅 2.5% 未听说过图数据库的概念。

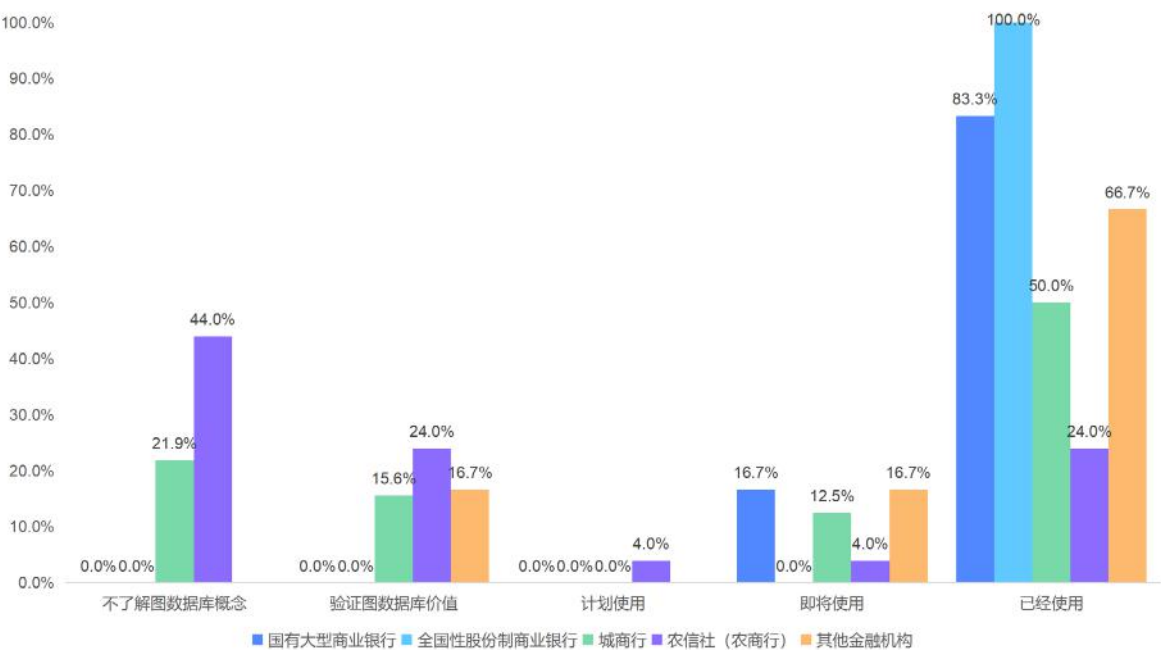


▲ 图7 金融机构图数据库使用现状

数据来源：《金融电子化》杂志社，2023年3月

当前，金融机构对于图数据库的认知存在两极分化现象。对图数据库技术本身的不了解及其场景落地价值的不清晰，成为制约金融机构引入图数据库产品的主要障碍。

对图数据库的使用进程主要集中在听说过但未使用过（20.3%），和已使用过一种以上的（51.9%），处于已了解、尚在探索或规划阶段（13.9%）（见图7）。



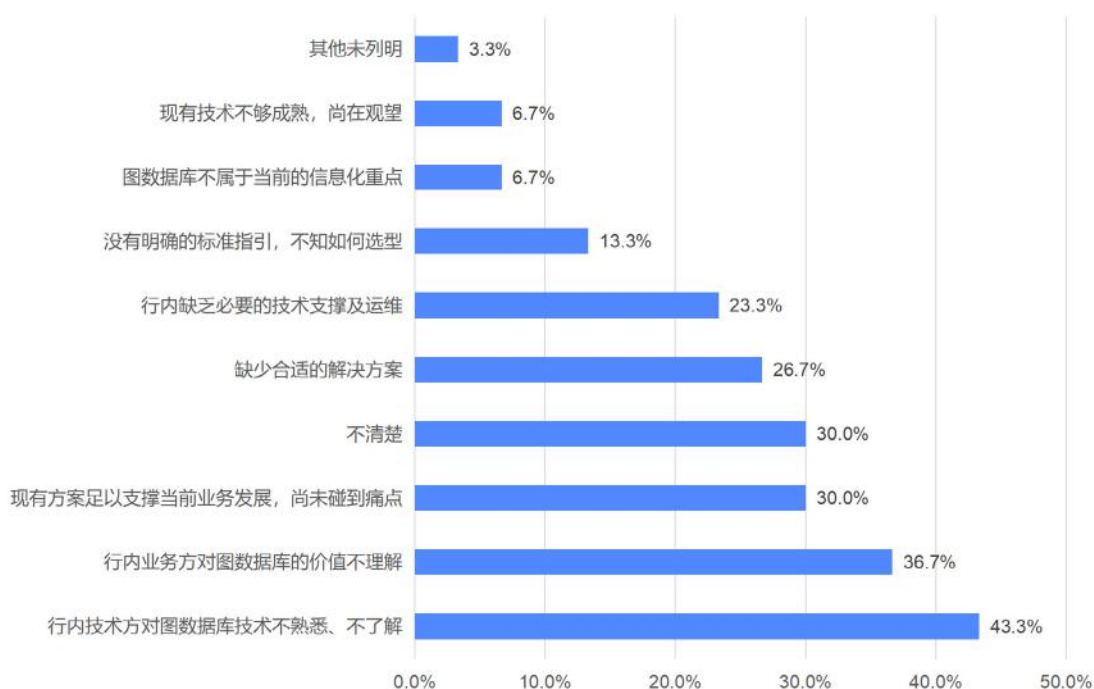
▲ 图8 各类金融机构图数据库使用现状对比

数据来源：《金融电子化》杂志社，2023年3月



其中，中小型金融机构对图数据库的认知更多处于初期阶段。68.0% 的农信社（农商行）、37.5% 的城商行尚在接触概念、探索价值，使用率尚未超过 50%，仅 5 家采购了商业图数据库。相比之下，国有大行、股份制银行则均对图数据技术有了深入了解，除一家机构仍处于 POC 测试阶段外，其他被调研机构均表已使用过一种及以上的图数据库，其中 66.7% 的国有大行、40.0% 的股份制银行已立项或采购了商业图数据库（见图 8）。

在对 30 家尚未使用图数据库且无采购计划的金融机构的调研发现，近 1/2 的被调研机构表示未使用图数据库产品的主要原因是对于图数据库技术（43.3%）及其业务价值（36.7%）不了解。近 1/3 的用户认为是尚未碰到技术痛点，1/4 的用户认为是缺乏合适的解决方案及技术支持能力。由此可见图数据库技术的市场普及推广仍然不足（见图 9）。



▲ 图 9 金融机构未应用图数据库技术的原因

数据来源：《金融电子化》杂志社，2023 年 3 月

在已经使用图数据库的 41 家金融机构中，图数据库建设进程在不同规模的金融机构间也存在较大差距。国有大行、股份制银行等对图数据库技术的整体投入及应用规模更大，充分体现了大行的技术领跑优势。

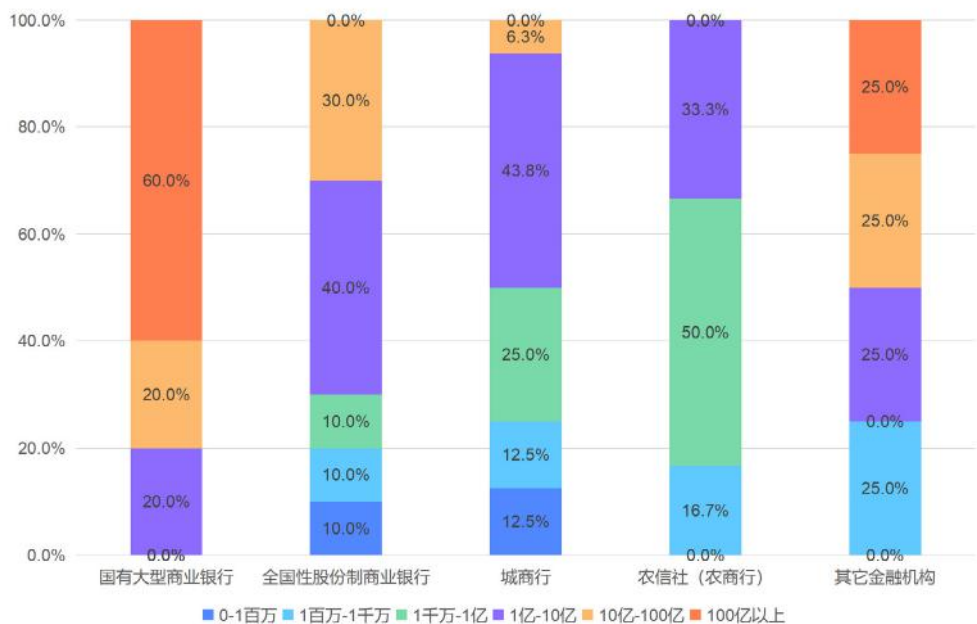
调研数据显示，60% 的国有大行的图数据规模超过 10TB，图规模超过 100 亿点边。40% 的国有大行图数据库部署规模达到 30 节点以上。50% 的股份制银行的应用则多在 1TB-10TB 数据、

1 亿 -100 亿点边图规模上，部署节点数也多在 4-10 节点的中型集群规模。城商行与农信社则更多集中在 2-3 节点，千万到亿级别的图规模应用。投资力度上，大行更倾向于投入 128 核之上的大算力服务器。（见图 10）。



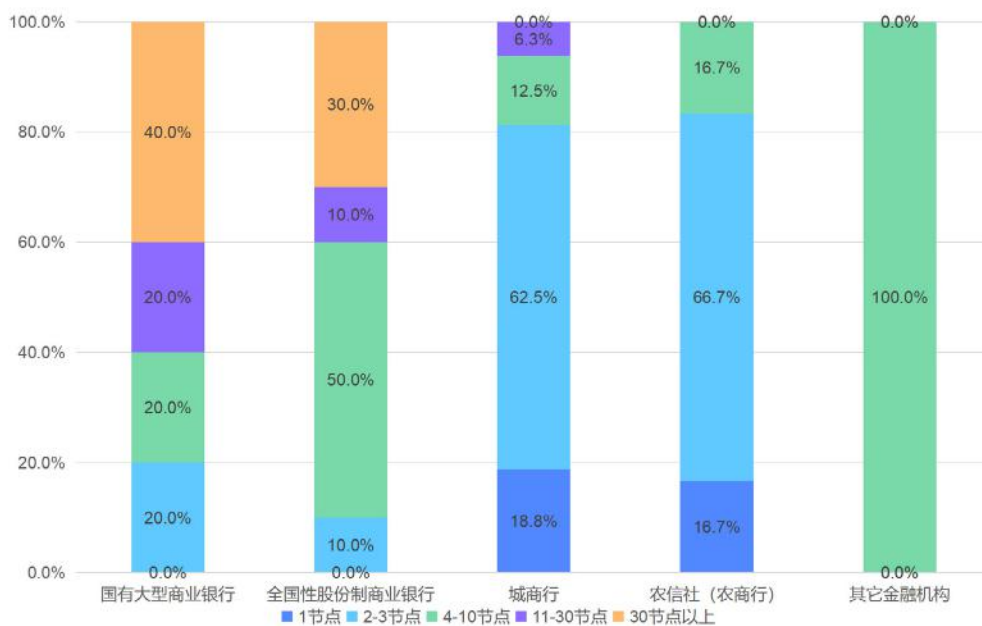
▲ 图 10-1 不同类型金融机构图数据库建设规模现状对比：存储规模

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 10-2 不同类型金融机构图数据库建设规模现状对比：图点边数量

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 10-3 不同类型金融机构图数据库建设规模现状对比：服务器节点数量

数据来源：《金融电子化》杂志社，2023 年 3 月

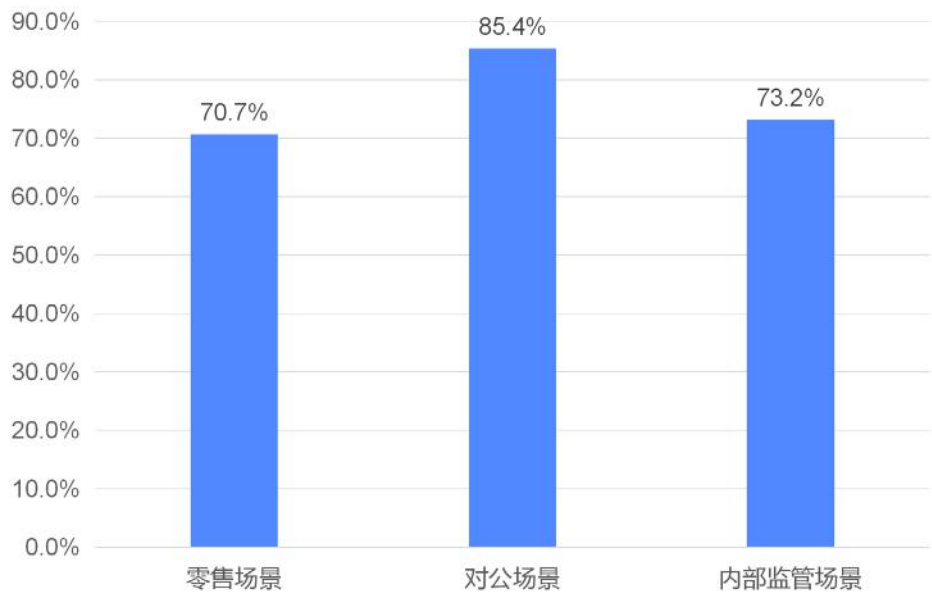


▲ 图 10-4 不同类型金融机构图数据库建设规模现状对比：服务器核数

数据来源：《金融电子化》杂志社，2023 年 3 月

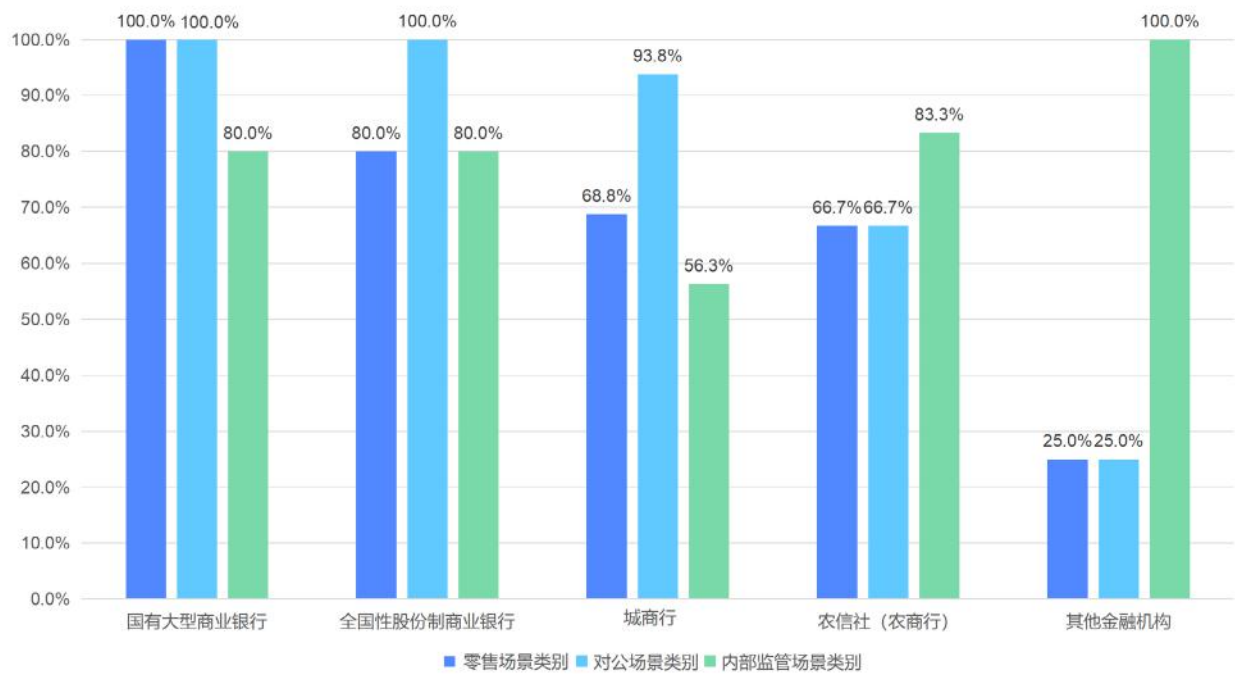
2. 图数据库应用场景多元发展

被调研的 41 家已部署图数据库的机构中，分别有 85.4%、73.2%、70.7% 的机构在对公、内部管理以及零售三类业务场景中应用了图数据库（见图 11）。



▲ 图 11 金融机构三大类场景中图数据库的应用率

数据来源：《金融电子化》杂志社，2023 年 3 月

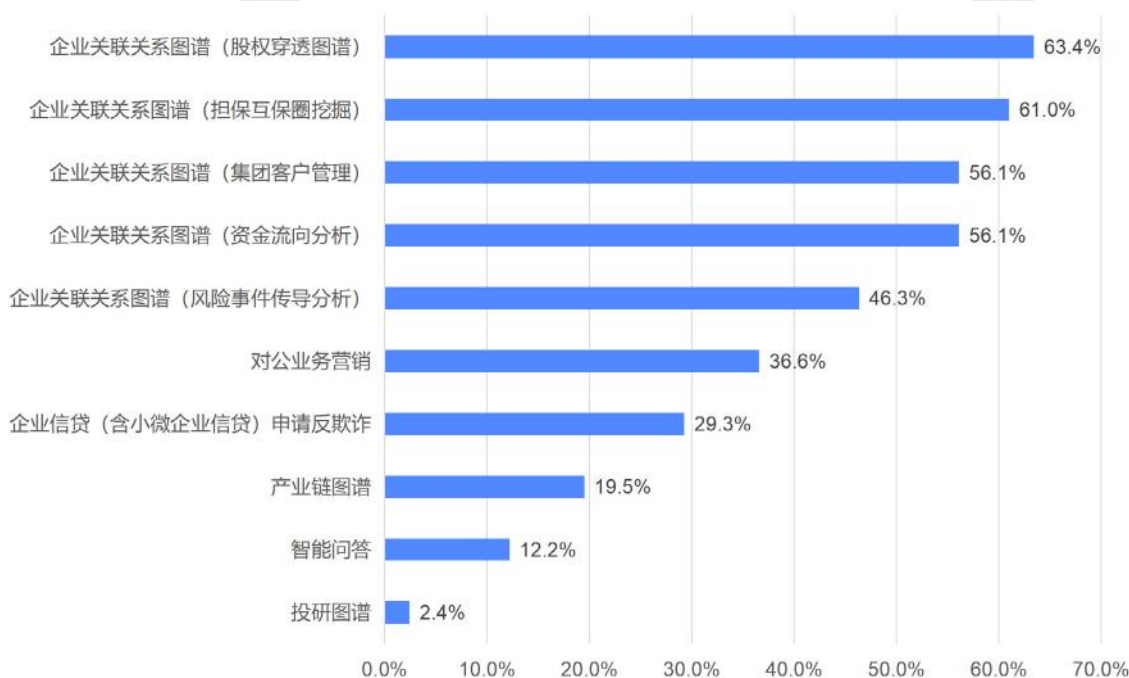


▲ 图 12 不同类型金融机构三大类场景中的图数据库应用率

数据来源：《金融电子化》杂志社，2023 年 3 月

总体上，国有大型商业银行在图技术应用上全面领跑其他金融机构，银行机构更多将图数据库应用于对公场景，保险和证券机构则更加专注于内部监管等相关场景（见图 12）。

在对公业务中，企业关系图谱是主要的应用场景，其中，股权穿透、担保互保圈挖掘为代表应用，超过 60% 的被调研机构表示已经落地相关应用（见图 13-1）。



▲ 图 13-1 对公业务中各场景的图数据库应用率

数据来源：《金融电子化》杂志社，2023 年 3 月

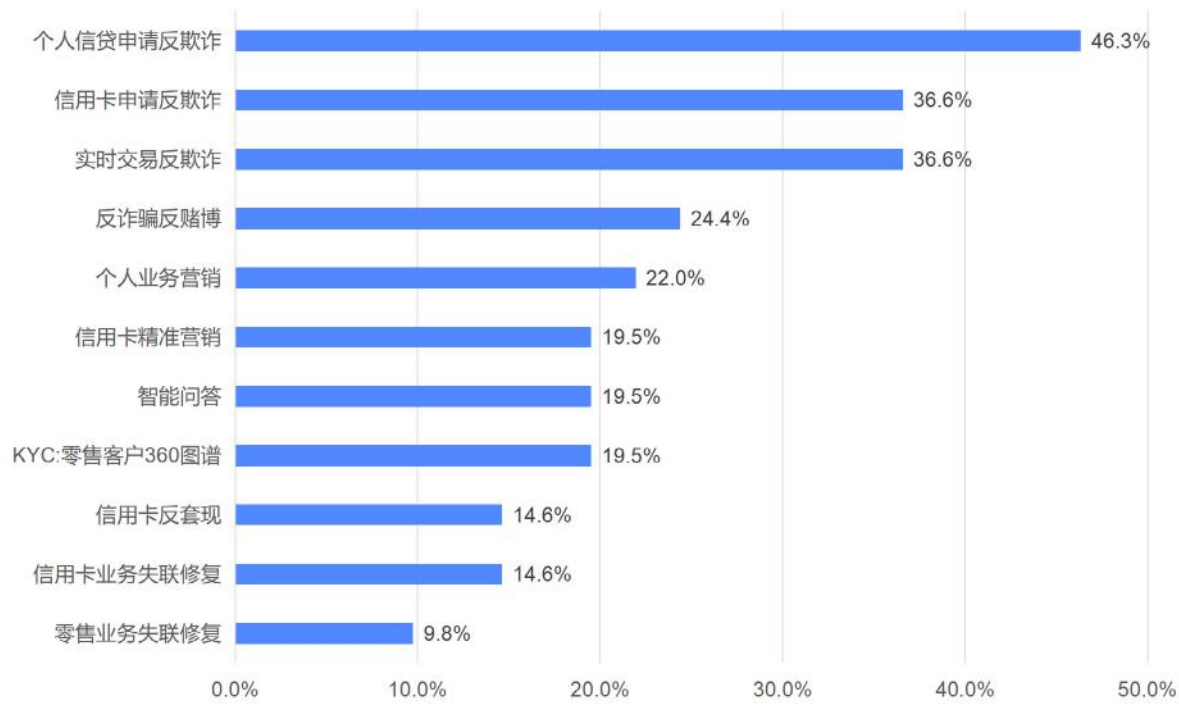
在零售业务中，申请反欺诈和交易反欺诈是主要应用场景，但整体图数据库应用普及率不及对公业务（见图 13-2）。

在内部监管业务中，近半数被调研机构已将图数据库应用于反洗钱场景，合规与审计是仅次于的应用场景（见图 13-3）。

不同机构间典型场景的应用情况也呈现出不同的分布特点。

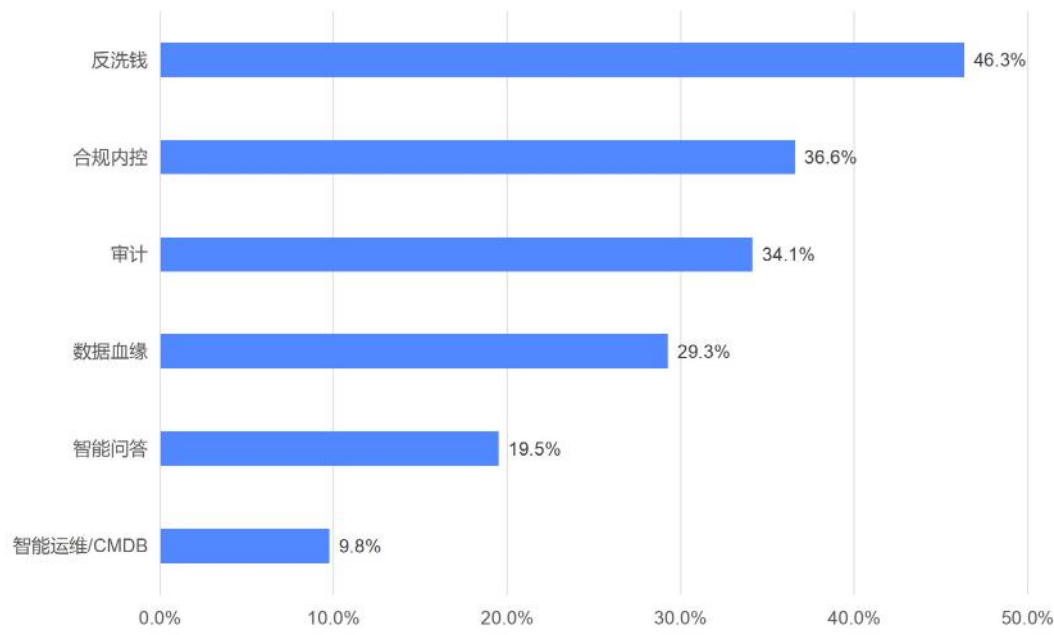
对公业务中（见图 14-1），不同机构间呈现出明显的梯度差异。除保险和证券机构外，图数据库在企业关联关系图谱相关场景（股权穿透图谱、担保互保圈挖掘、集团客户管理、资金流向）中应用率高，国有大行 100% 覆盖；在产业链图谱、投研图谱场景中，各金融机构普遍应用率低。此外，对公业务中图数据库在营销场景的应用率整体远低于风险场景。在国家政策大力推广的小微信贷业务中，图数据库的应用率也低于传统对公业务，未来的市场发展空间较大。

在零售业务中（见图 14-2），不同类型机构应用情况差异较大。国有大行在个人信贷反欺诈



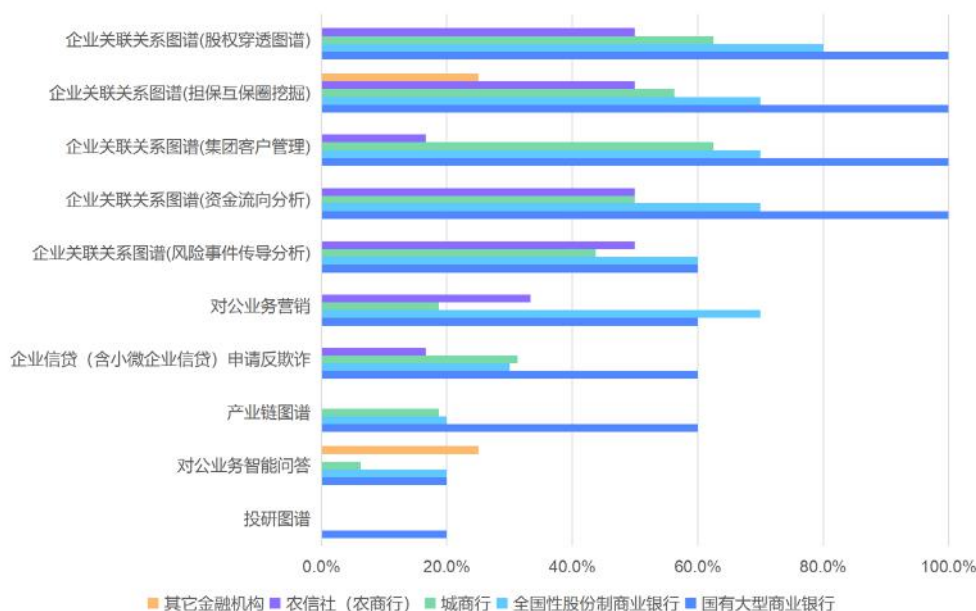
▲ 图 13-2 零售业务中各场景的图数据库应用率

数据来源：《金融电子化》杂志社，2023 年 3 月



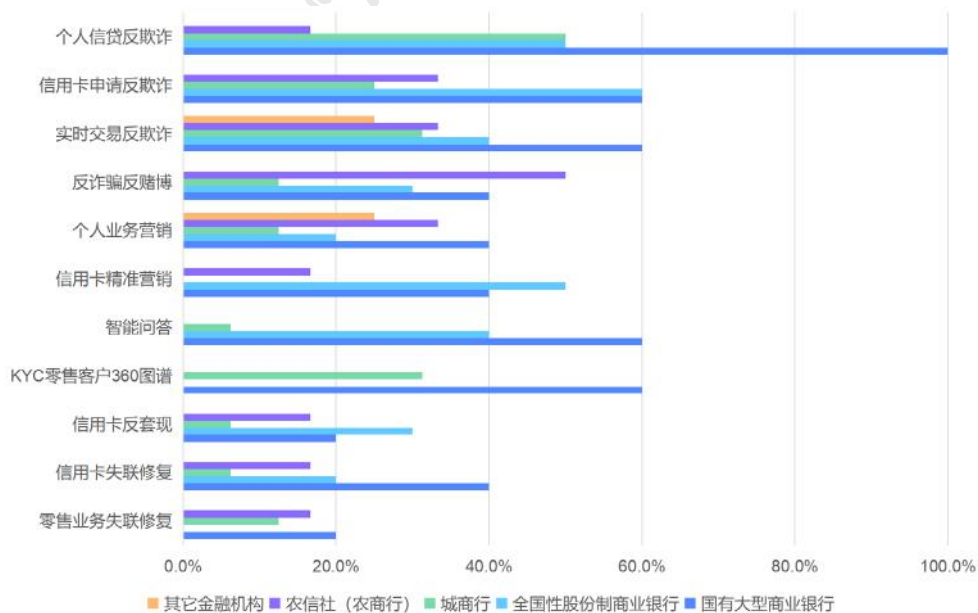
▲ 图 13-3 内部监管各场景的图数据库应用率

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 14-1 对公业务：不同类型金融机构各场景图数据库应用率对比

数据来源：《金融电子化》杂志社，2023 年 3 月

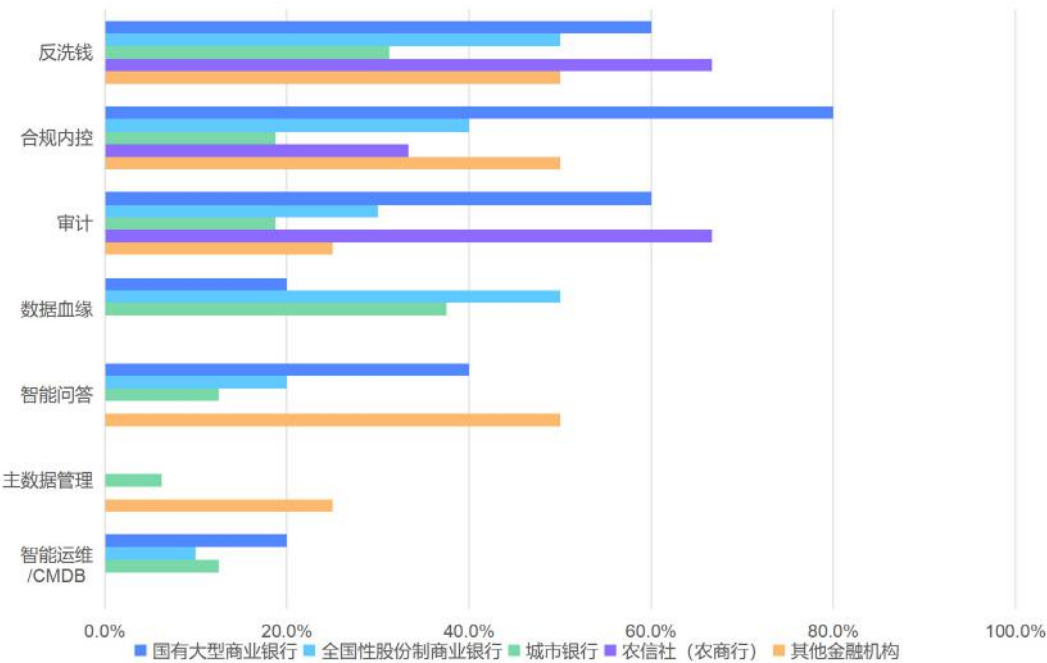


▲ 图 14-2 零售业务：不同类型金融机构各场景图数据库应用率对比

数据来源：《金融电子化》杂志社，2023 年 3 月

场景下的应用率达到 100%，远超其他类型机构的场景应用率。相较之下，股份制银行更多投入在信用卡相关场景（申请反欺诈、精准营销、反套现）。农信社（农商行）则在反诈骗、反赌博场景方面投入高于国有大行、股份制银行及城商行。值得一提的是，被调研的股份制银行及农信





▲ 图 14-3 内部监管：不同类型金融机构各场景图数据库应用率对比

数据来源：《金融电子化》杂志社，2023 年 3 月

社（农商行）都未应用图数据库技术刻画零售客户 360 图谱，此场景或为零售业务图数据库应用的一个增长点。

对公及零售业务中，国有大行、股份制银行的图数据库应用率更广。内部监管业务中（见图 14-3），农信社（农商行）的图数据库应用率更高。审计及反洗钱场景中，农信社（农商行）应用率超过国有大行、股份制银行及城商行。

对保险和证券机构而言，图数据库则主要应用于内部监管相关场景。反洗钱、智能问答、合规内控等场景应用率达到 50%。数据血缘、主数据管理、智能运维等提升银行 IT 部门自身运营、管理效率的场景，虽然当前整体应用率较低，但未来的市场增长空间大。

### 三、金融业图数据库选型关注重点

选择图数据库产品时需要从性能、安全可控性、业务支撑度、简单易用性等多个维度分析图数据库产品间的核心差异，应综合考虑图数据库大数据量下的查询性能、大规模部署下的稳定性、代码安全性、是否自主可控、是否具有解决方案能力、能够支撑业务逻辑的定制化开发、技术运维工具是否完备完善等多个方面。我们将这些技术要素按照性能、业务支撑度、安全可控、简单易运维四个维度进行分类（见表 1）。

针对 49 份已立项、正在 POC、已使用图数据库的被调研机构有效问卷的排序，统计各类要



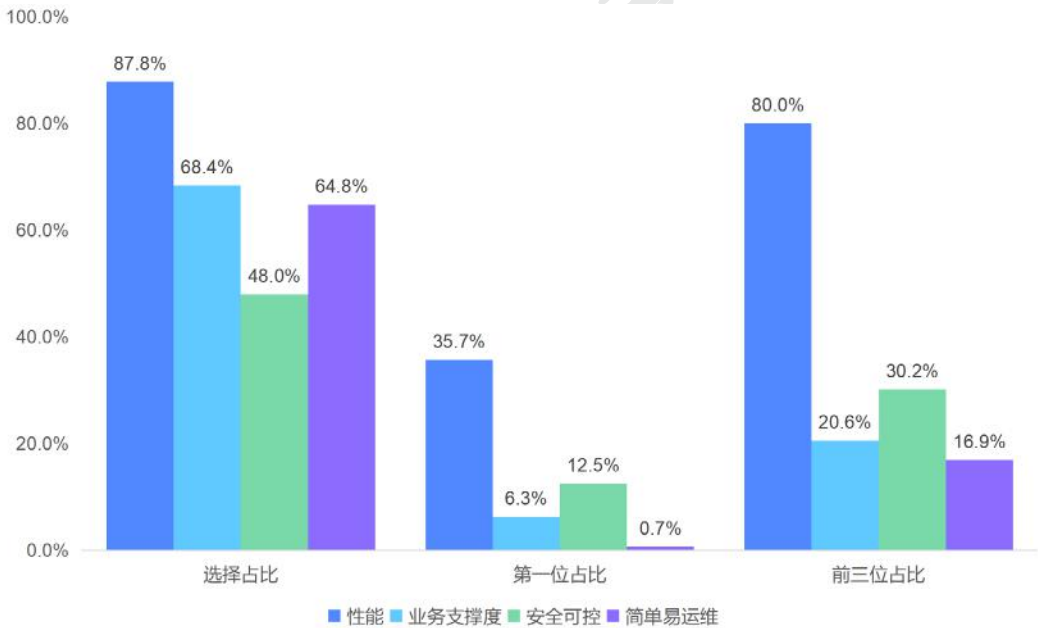
表 1 图数据库技术评估因素及其分类

性能	大数据量下的查询性能，随业务增长性能的稳定性
	支持原生分布式架构，便于大规模部署，保障业务的连续性
业务支撑度	图数据库供应商是否拥有解决方案提供能力
	图分析功能是否丰富
	对业务部门是否简单易用
安全可控	应用生态是否完善，与多少解决方案商已完成适配
	是否闭源，代码安全性能否保障、知识产权风险能否规避
	是否 100% 自主可控
简单易运维	是否具备丰富的开发接口
	是否具备完备的培训体系
	是否具备完备的售后服务体系
	对技术部门是否简单易运维

素被选择的数量占比、被选择的顺序在首位的占比、被选择的顺序在前三位的占比，得出四大维度在金融机构评价体系中的重要程度（见图 15）。

1. 性能表现成为金融机构关注焦点

随着数据规模急剧增长，存储处理的数据类型、数据维度多样化、复杂化，金融机构首要关注的是图数据库的大数据处理性能。87.8% 的被调研机构认为性能表现是重要的产品技术价值，



▲ 图 15 四个图数据库评估维度在金融机构采购决策中的重要性统计

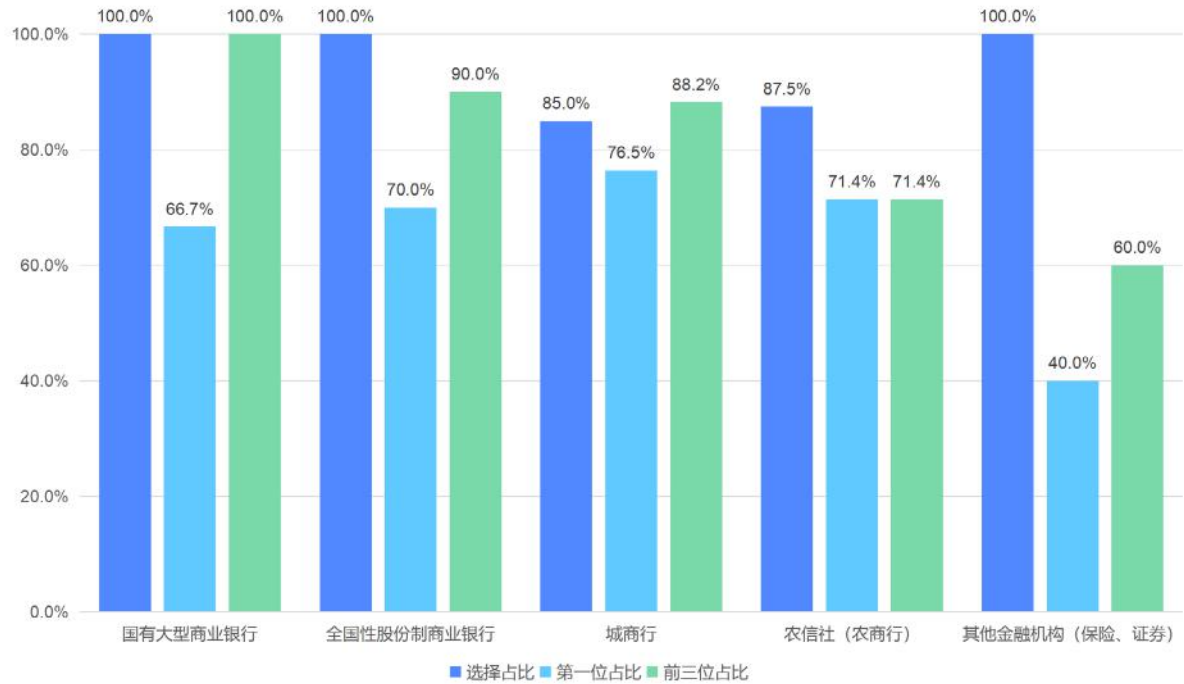
数据来源：《金融电子化》杂志社，2023 年 3 月

其中 35.7% 的机构将大数据量下的查询性能或是否支持原生分布式架构的性能相关指标重要程度排在首位。80.0% 的被调研机构将大数据下的查询性能或是否支持原生分布式架构的性能相关指标重要度放在所有指标的前三位。

大数据量下的查询性能

随着金融机构存储、处理的数据量（点数）增大，数据间的复杂关联关系（边数）将呈指数级增长，这对图数据库存储、处理大数据的能力提出了很高的要求。91.8% 的机构认为图数据库在大数据量下的查询性能是重要指标，其中 68.9% 将其看作首要关注指标。

国有大行、股份制银行、保险及证券机构对于性能的关注度更高，100% 的被调研机构认为大数据查询性能重要，但只有半数机构将其重要性排在第一；股份制银行、城商行及农信社（农商行）并不都认为该指标重要，认为该指标重要的机构更多将其排在了相对更重要的位置，超 70% 的机构将其排在了首位（见图 16）。



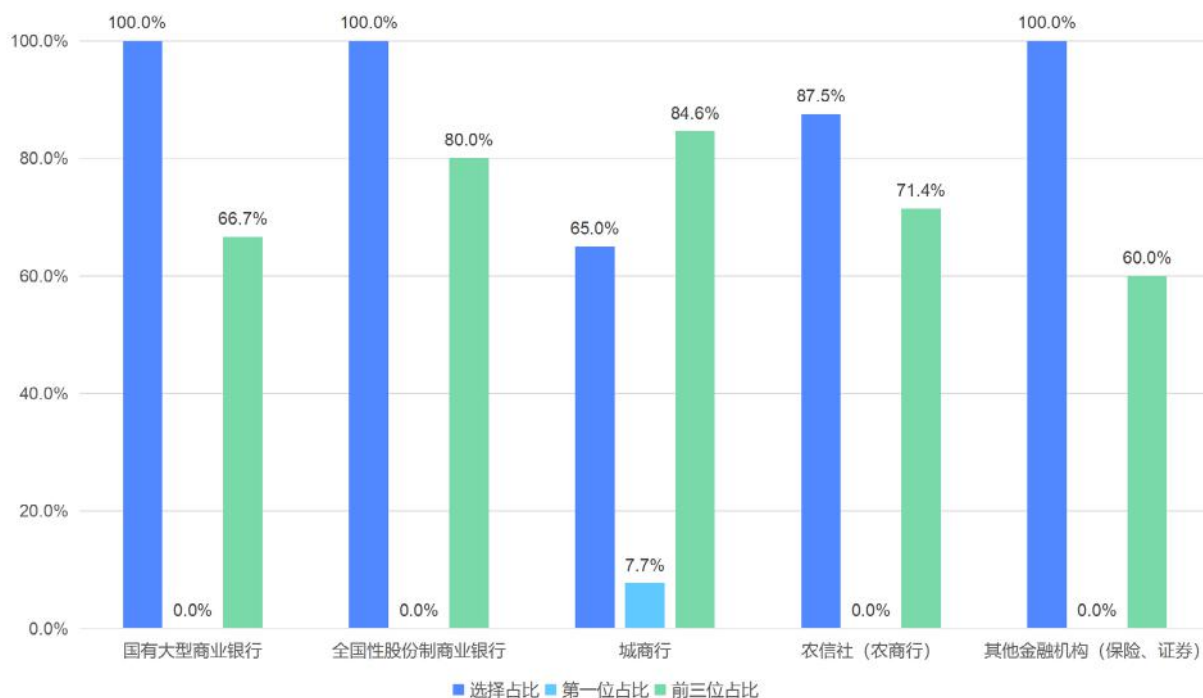
▲ 图 16 不同金融机构对大数据量下查询性能的重要性评估

数据来源：《金融电子化》杂志社，2023 年 3 月

是否支持原生分布式架构

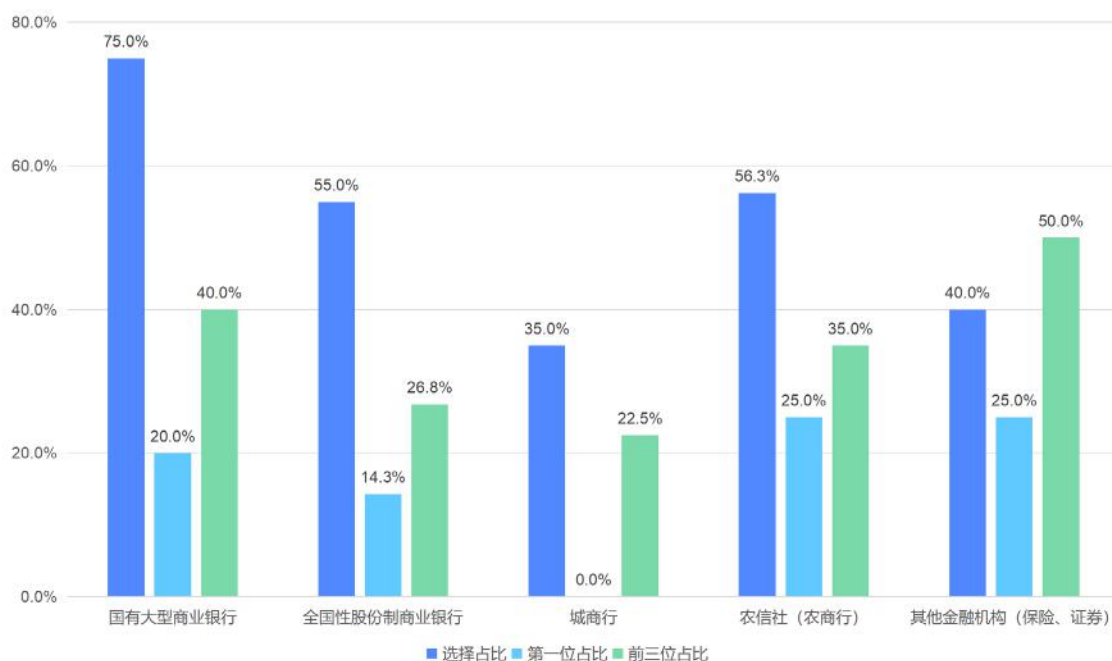
原生分布式架构能够保证图数据库的横向扩展力与分布式的查询性能。75.6% 的被调研机构将其放在所有技术价值指标的前三位。

与大数据的查询性能相似，在被调研的国有大行、股份制银行、保险及证券机构中，100%认为原生分布式架构重要，但极少机构将其作为首要考虑因素（见图 17）。



▲ 图 17 不同类型金融机构对原生分布式架构的重要性评估

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 18 不同类型金融机构对安全可控的重要性评估

数据来源：《金融电子化》杂志社，2023 年 3 月

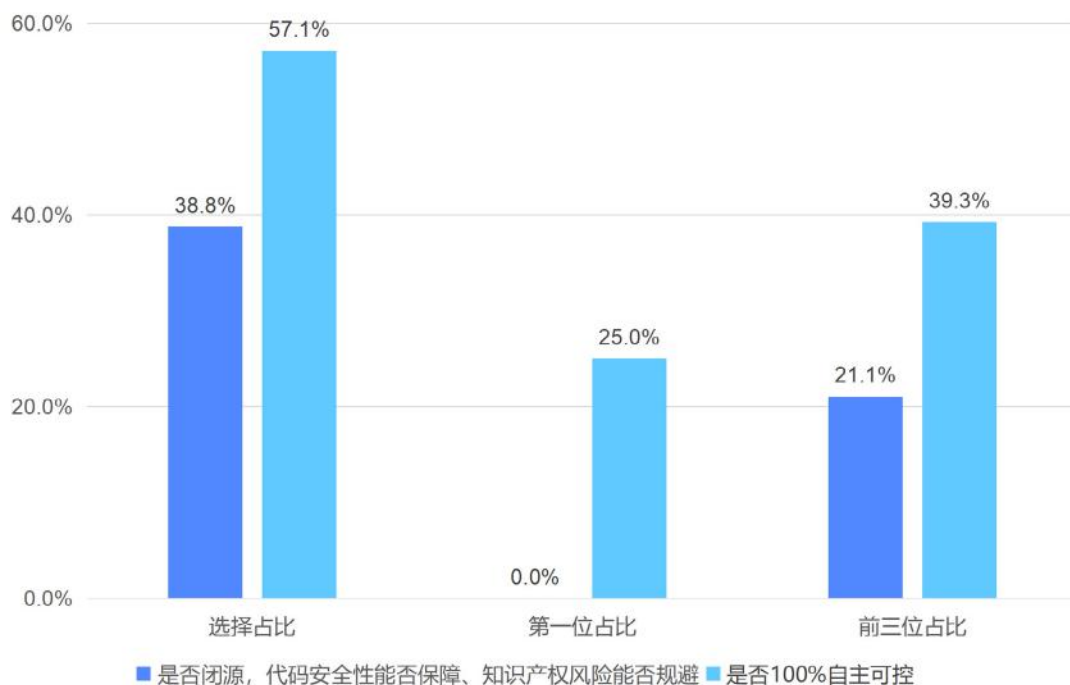
相比其他类型的银行，国有大行更看重图数据库产品其他方面的能力，尤其是安全性。

## 2. 安全可控是金融机构的重要考虑因素

数据库技术的安全可控主要包含闭源和自主可控两方面的因素。48% 的被调研机构考虑安全可控的相关因素，30.2% 将其放在了所有因素的前三位。尽管超过 60% 的被调研机构认为“业务支撑度”“简单易运维”重要，但将这些因素排在前三位的比例非常低。

不同类型机构中，国有大行普遍对于安全可控的关注度更高，75% 的国有大行认为该因素重要；更多的农信社（农商行）、保险证券机构将其作为首要考虑因素。（见图 18）

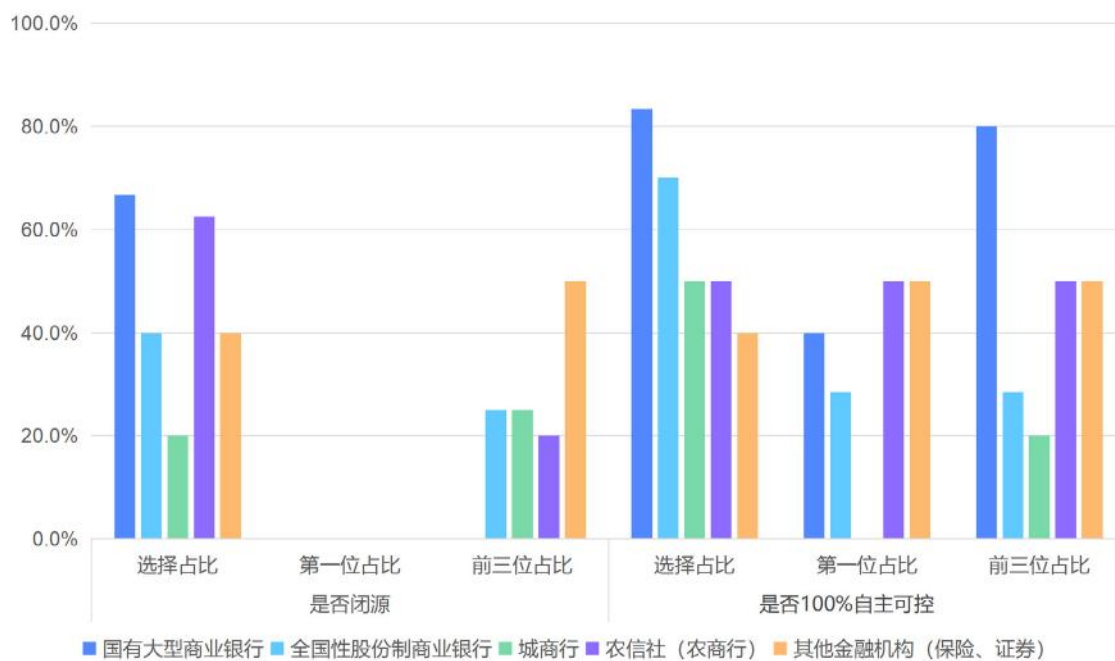
开源产品相比闭源产品有更高的知识产权风险、易受攻击的安全隐患，存在卡脖子的风险。开源社区项目的数据统计显示，没有大厂持续投入的开源项目商业服务持续性与稳定性不佳，造成产品的总体使用成本更高。统计发现，近 40% 的被调研机构认为闭源为重要因素（见图 19），其中，50% 的保险、证券机构将其重要性排在所有考虑因素的前三。（见图 20）。



▲ 图 19 图数据库在安全可控维度下的性能比较

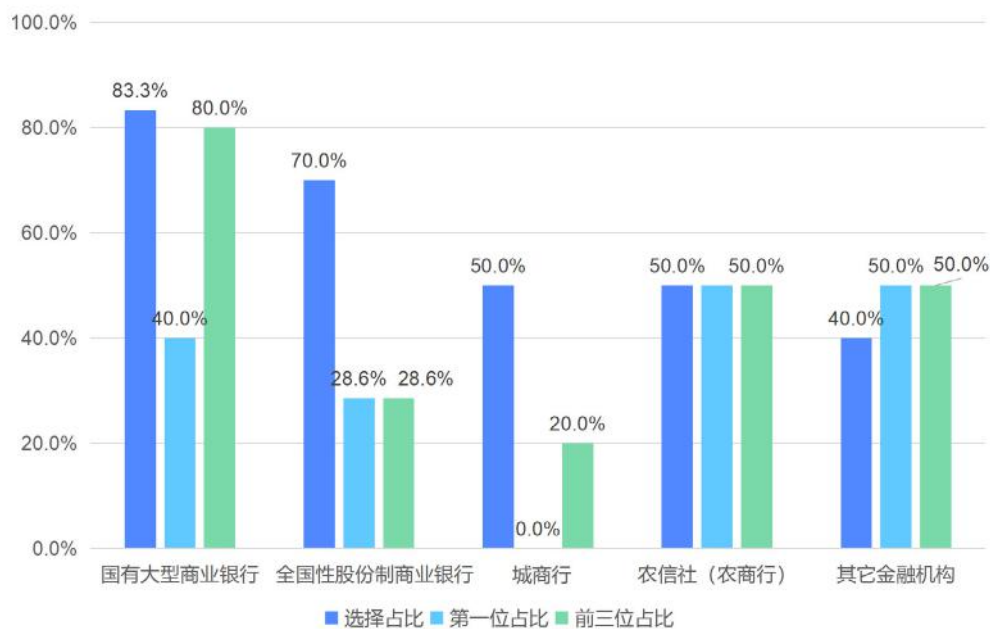
数据来源：《金融电子化》杂志社，2023 年 3 月

自主可控体现在两个方面，一是产品的知识产权及核心代码的开发人员都属于中国企业，二是企业的实控人及资本结构为内资。50% 以上的金融机构认为自主可控重要，83.3% 的国有大行、70.0% 的股份制银行将自主可控作为重要考虑因素。在认为自主可控重要的机构中，40.0% 的国



▲ 图 20 不同类型金融机构对安全可控的评估对比

数据来源：《金融电子化》杂志社，2023 年 3 月



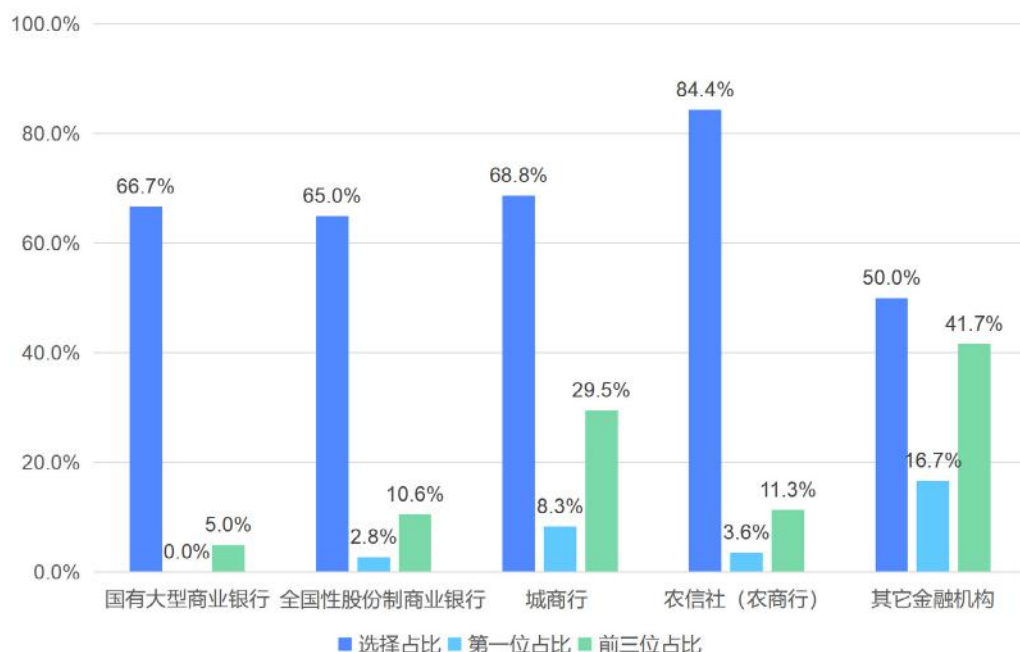
▲ 图 21 不同类型金融机构对是否 100% 自主可控的重要性评估对比

数据来源：《金融电子化》杂志社，2023 年 3 月

有大行将其列为首要考虑因素，80.0% 将其排在了前三位。

### 3. 业务支撑能力是普遍关注点

对于图数据库厂商对业务的支撑能力，金融机构同样表现出了较高的关注度。68.4% 的被调研机构认为业务支撑能力十分重要，20.6% 的被调研机构将其列在所有技术因素的前三位，84.4% 的农信社（农商行）最关注业务支撑能力（见图 22）。



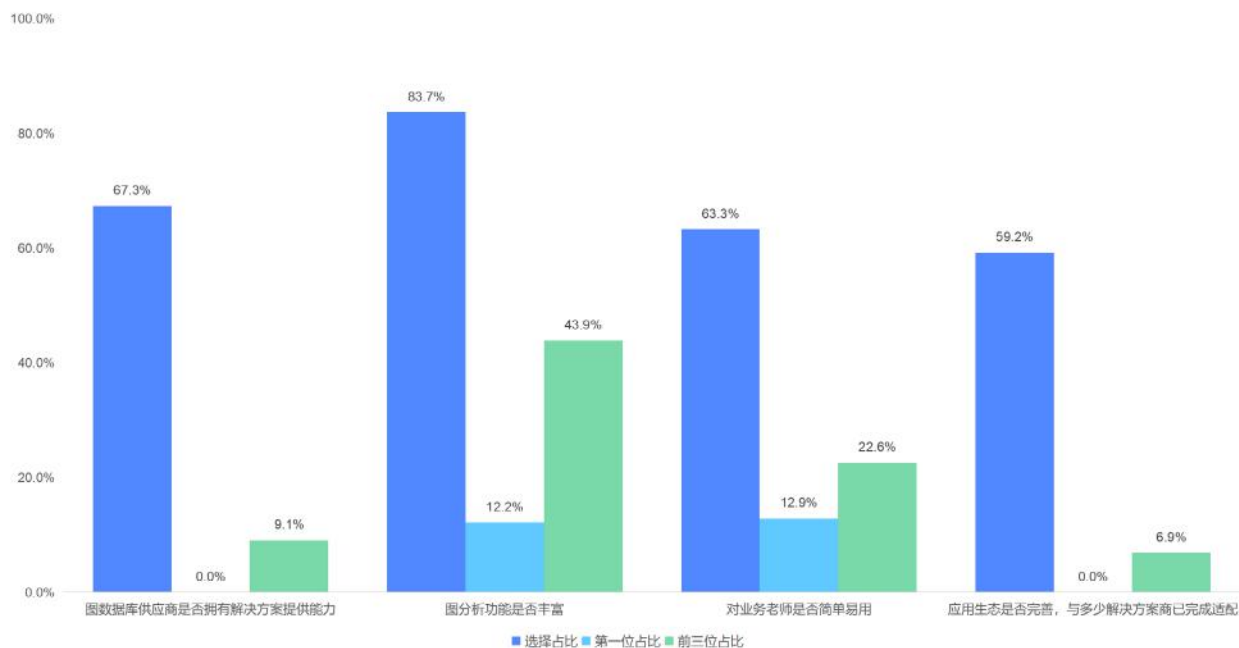
▲ 图 22 不同类型金融机构对业务支撑能力的重要性评估对比

数据来源：《金融电子化》杂志社，2023 年 3 月

图数据库的业务支撑能力主要体现在以下四方面：图数据库的供应商是否有解决方案供应能力；图数据库是否具备丰富的图分析功能；对业务人员是否简单易用；应用生态是否完善。关注图数据库厂商业务支持能力的机构中，83.7% 的金融机构关注图数据库是否具备丰富的图分析能力，43.9% 将其排在了所有考虑因素的前三位。值得关注的是，67.3% 的金融机构期望图数据库供应商具备提供完整解决方案的能力（见图 23）。这些需求意味着，不同于传统关系型数据库市场，图数据库产品需要在数据存储、查询能力之外，建设更完备的数据分析能力及方案支撑能力。

### 4. 简单易运维受行业青睐

图数据库作为一种新兴技术，学习门槛高、市场人才少，因此，各类金融机构普遍对图数据库的便捷运维性提出了要求。64.8% 的被调研机构认为简单易运维的相关特点是选择图数据库的重要考量因素。16.9% 的被调研机构将该类别的相关因素排在了所有因素的前三位。

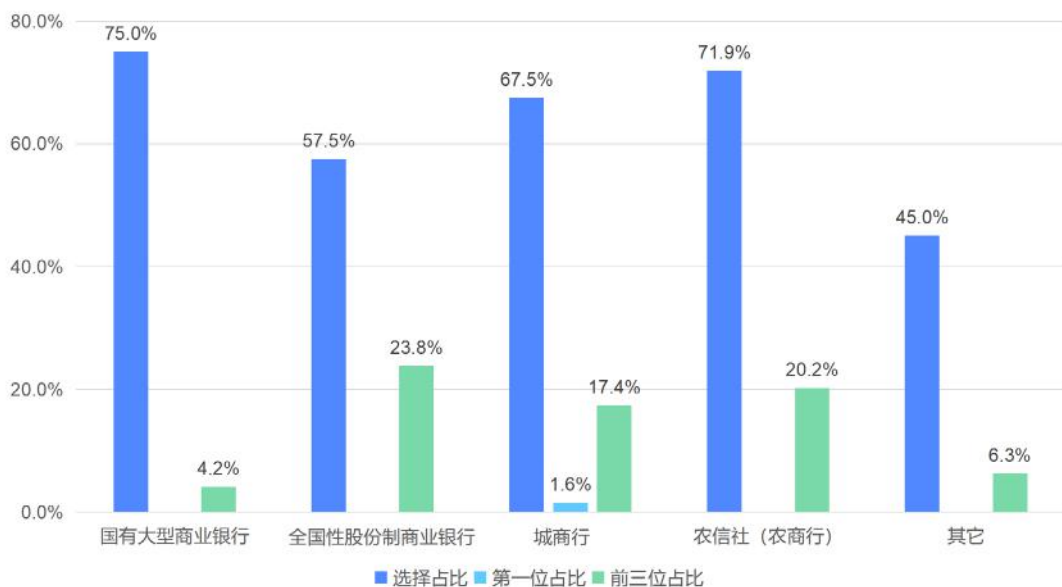


▲ 图 23 业务支撑能力各技术要素重要性评估对比

数据来源：《金融电子化》杂志社，2023 年 3 月

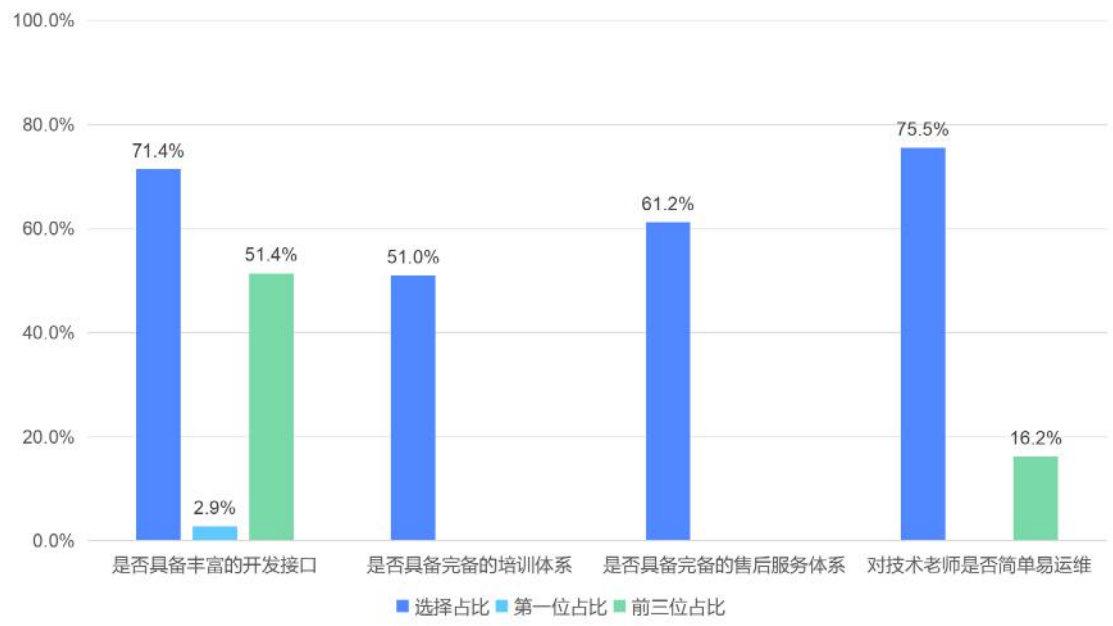
75.0% 的国有大行、71.9% 的农信社（农商行）、67.5% 的城商行和 57.5% 的股份制银行关注简单易运维（见图 24）。

图数据库的易运维程度取决于产品开发接口的丰富程度、培训体系与售后服务体系的完备程度等。这其中，是否具备丰富的开发接口成为最重要的考虑因素。因为金融机构业务复杂多样，



▲ 图 24 不同类型金融机构对简单易运维的重要性评估

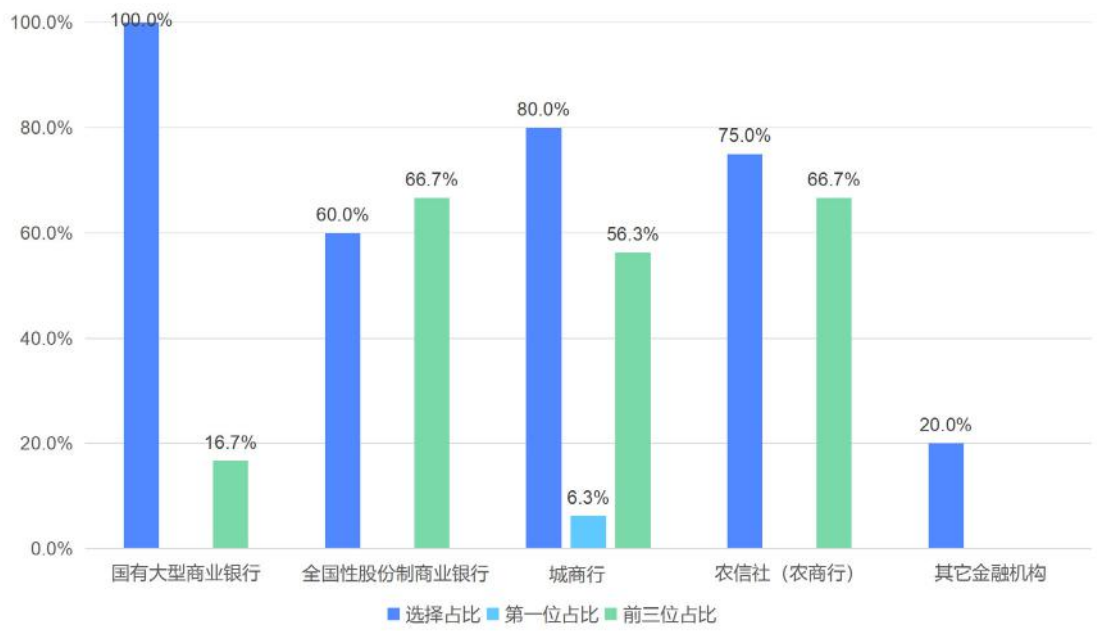
数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 25 简易运维各技术要素重要性评估对比

数据来源：《金融电子化》杂志社，2023 年 3 月

丰富的开发接口能够帮助金融机构简易地实现业务的深度定制化开发。51.4% 的金融机构将其排在重要性前三位（见图 25），少数城商行甚至将其评为最重要的因素（见图 26）。



▲ 图 26 不同类型金融机构对丰富开发接口要素的重要性评估

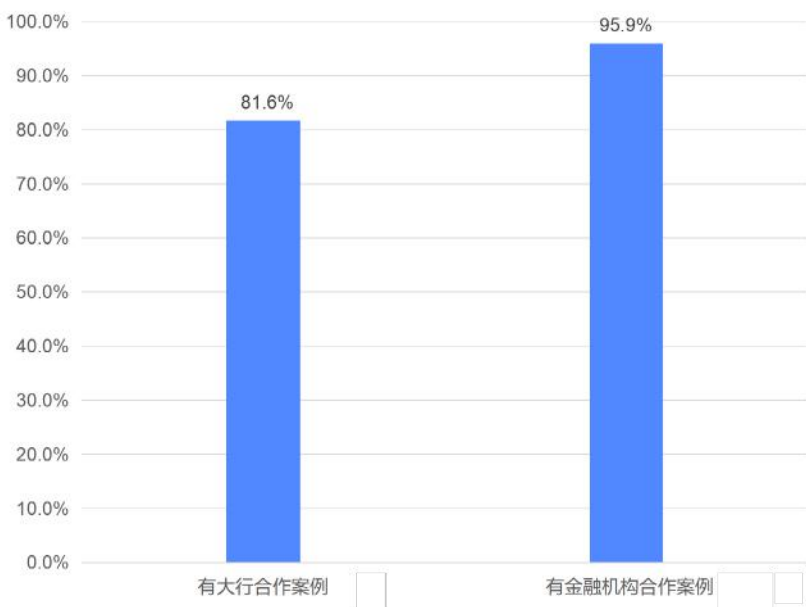
数据来源：《金融电子化》杂志社，2023 年 3 月



## 5. 用户普遍关注企业在行业的应用能力

图数据库技术正处在从小众市场逐步进入大众市场。不同于早期用户看重产品的技术性能，大众市场的用户更看重产品的行业实践案例。

因此，除了产品的技术能力，图数据库厂商的商业化案例也成为影响金融机构决策的关键因素。对于图数据库厂商而言，丰富的落地案例意味着拥有成熟的产品和稳定的系统。调研结果显示，95.9% 的被调研机构表示希望厂商有金融机构合作案例，其中 81.6% 看重是否有国有大行的案例（见图 27）。



▲ 图 27 金融机构对图数据库企业在行业的应用能力要求

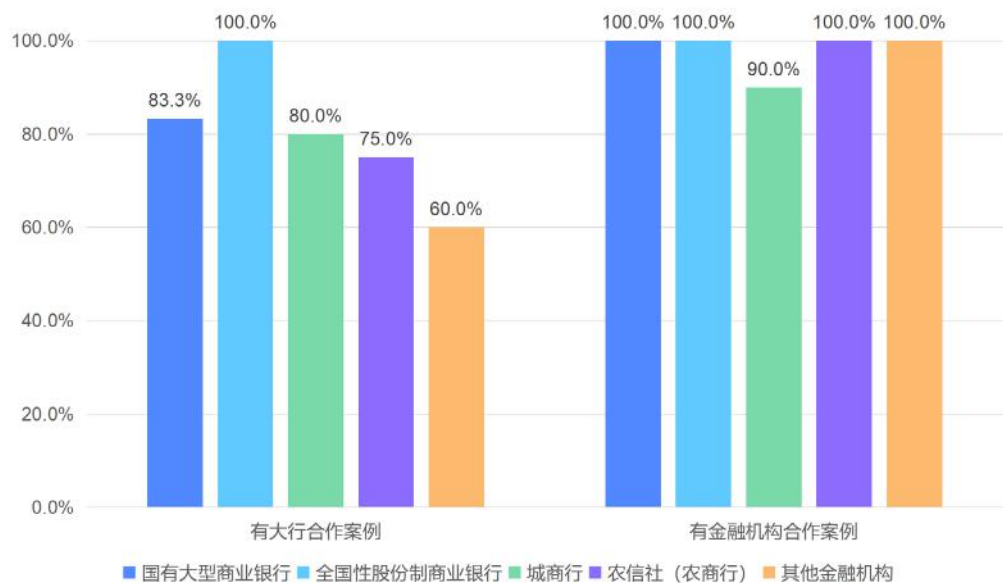
数据来源：《金融电子化》杂志社，2023 年 3 月

国有大行具有丰富的业务场景、复杂的技术架构、数据及业务处理要求高，能很好地促进图数据库厂商产品迭代。除保险、证券机构外，各银行机构普遍看重和有国有大行案例。100% 被调研的股份制银行将其作为重要考虑因素（见图 28）。

## 四、金融业图数据库建设难点

### 1. 人才紧缺、业务场景价值有待验证是建设的主要瓶颈

由于图数据库属于新兴领域，学习门槛高，应用难度大，亟需专业的图技术复合型人才。图技术专业人才的稀缺已经成为制约图数据库金融行业应用的重要因素，78.5% 的被调研机构表示缺乏相关的专业技术人才。已经使用了图数据库的金融机构更明显，高达 87.8% 的机构认为专业

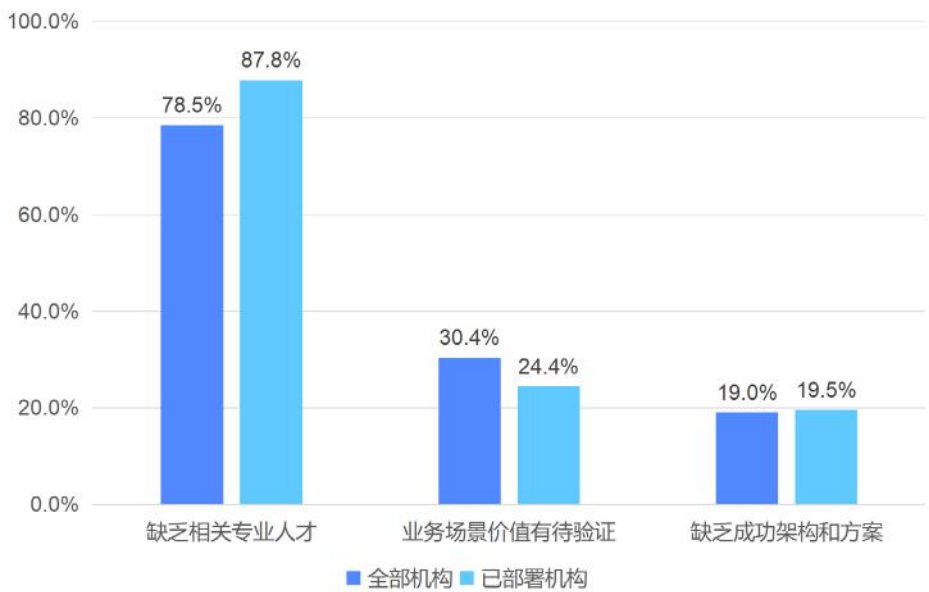


▲ 图 28 不同类型金融机构对企业在行业的应用能力要求对比

数据来源：《金融电子化》杂志社，2023 年 3 月

人才的缺乏是重要痛点。

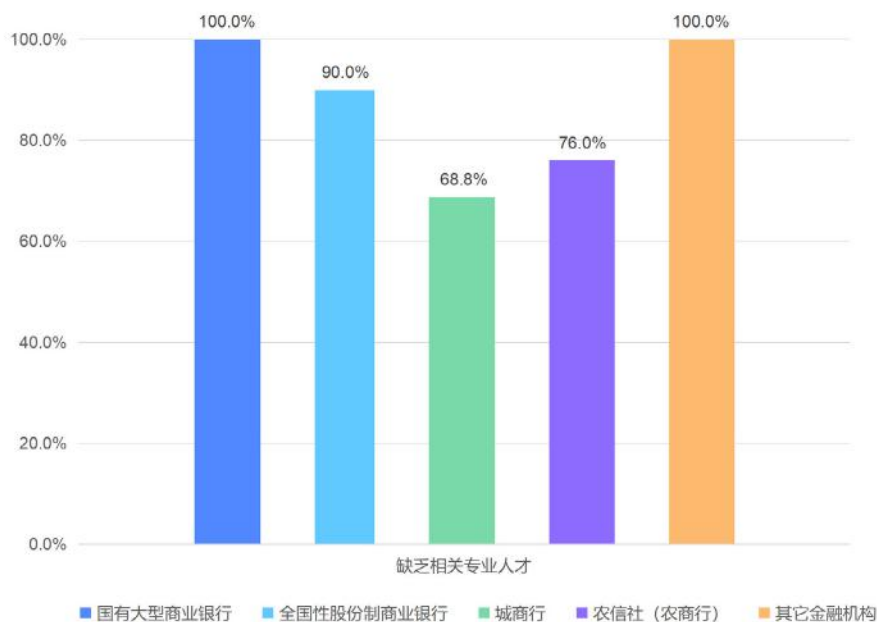
人才的缺乏间接导致了图数据库业务价值难以发挥，缺乏相关专业人才意味着图技术在实际应用过程中，金融机构缺乏系统的知识体系，不知道哪里用、怎么用。即使是已使用图数据库的金融机构，依然有 24.4% 的被调研机构表示图数据库在其业务场景下的价值有待验证，19.5% 的被调研机构表示他们缺乏成功的架构和方案（见图 29）。



▲ 图 29 金融机构应用图数据库的主要难点

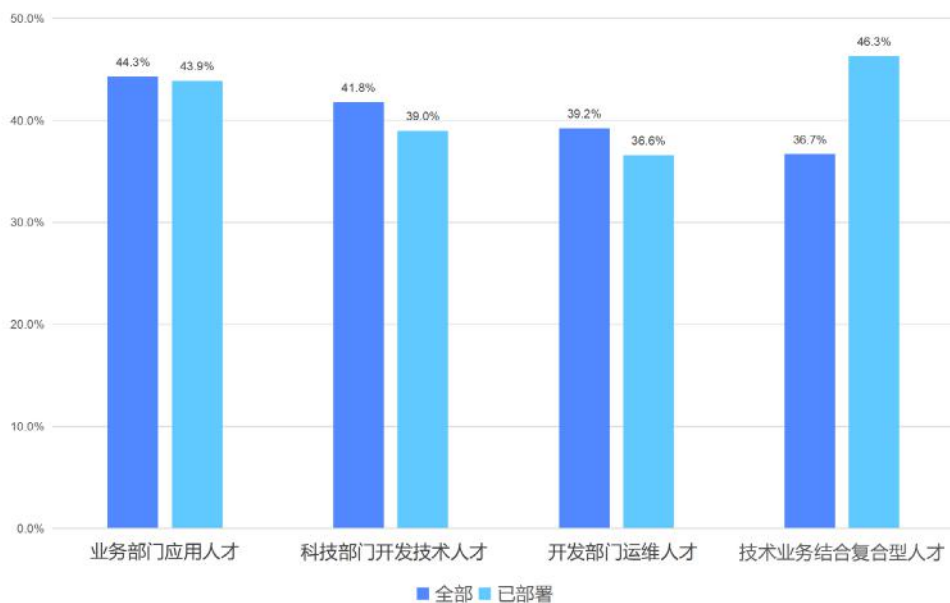
数据来源：《金融电子化》杂志社，2023 年 3 月

围绕不同金融机构的人才需求情况,100%的国有大行、保险和证券机构均表示缺乏相关人才(见图30)。按技术人才的类型划分,金融行业最缺乏的是技术业务结合的复合型人才(见图31)。



▲ 图30 不同类型金融机构对图技术人才的需求对比

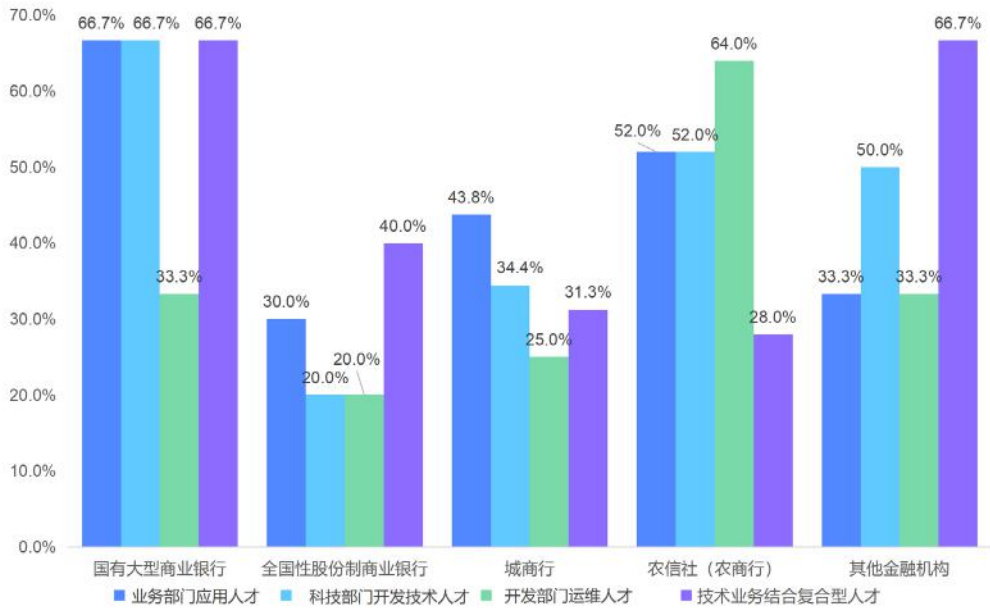
数据来源：《金融电子化》杂志社，2023年3月



▲ 图31 金融机构对各类图数据库人才的需求对比

数据来源：《金融电子化》杂志社，2023年3月

国有大行作为图数据库技术的深度用户，对业务、技术，以及复合型人才均需求较高，农信社（农商行）等机构相对缺乏图数据库的运维人才（见图 32）。



▲ 图 32 不同类型金融机构对各类图数据库人才的需求对比

数据来源：《金融电子化》杂志社，2023 年 3 月

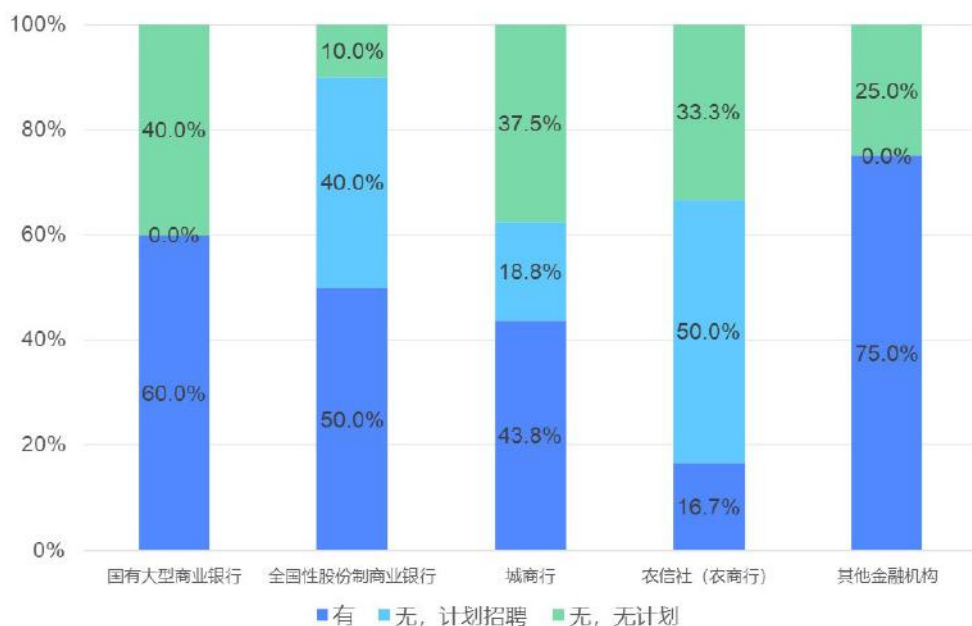
不同类型的金融机构在图数据库人才梯队建设上也存在明显差异。图数据库应用更成熟的国有大行人人才建设更加完善，股份制银行的人才需求最为迫切（见图 33-1）。79 家被调研机构中，66.7% 的国有大行、50% 的股份制银行与保险证券公司均表示已经招聘了相关人才。56.3% 城商



▲ 图 33-1 被调研机构的人才现状及招聘计划

数据来源：《金融电子化》杂志社，2023 年 3 月

行、80% 的农信社（农商行）无人才招聘计划，原因是该群体对图数据库技术理解、应用程度不高。已经部署了图数据库的城商行与农信社（农商行）的人才建设需求显著高于未部署的同类机构，但整体的人才建设水平依然不及全国性银行（见图 33-2）。



▲ 图 33-2 已部署图数据库的金融机构的人才现状及招聘计划

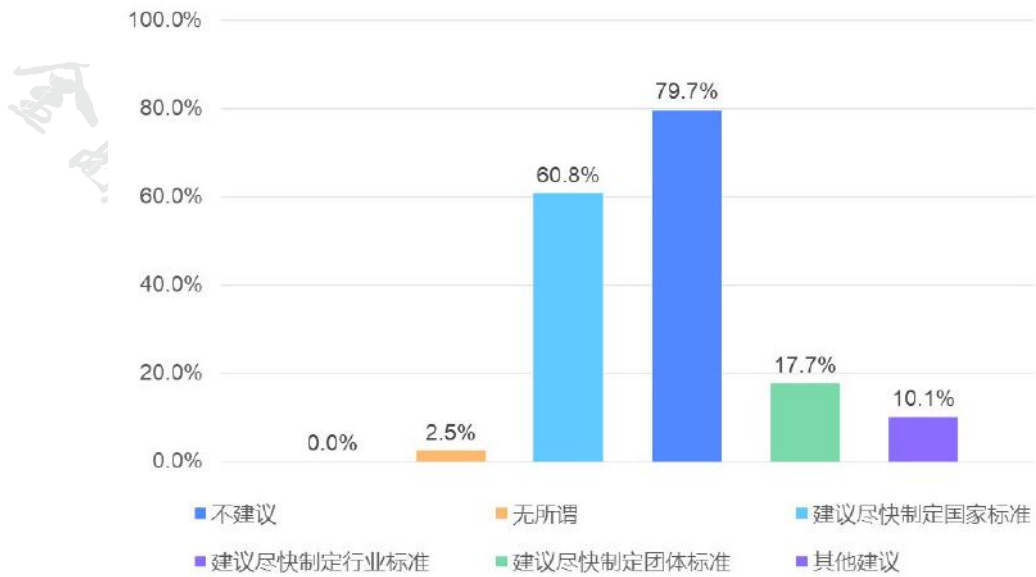
数据来源：《金融电子化》杂志社，2023 年 3 月

## 2. 亟待建立与推广图数据库行业标准

当前，图数据库技术正在快速发展，但仍处于早期发展阶段，缺乏行业标准和业内共识。缺失国家、行业、团体标准，制约了图数据库的应用以及商业化落地进程，不利于图数据库行业的发展。调研结果显示，金融机构迫切需要相关部门制订图数据库行业标准，60.8% 的被调研机构建议制订国家标准，79.7% 的被调研机构希望加快制订行业标准（见图 34）。

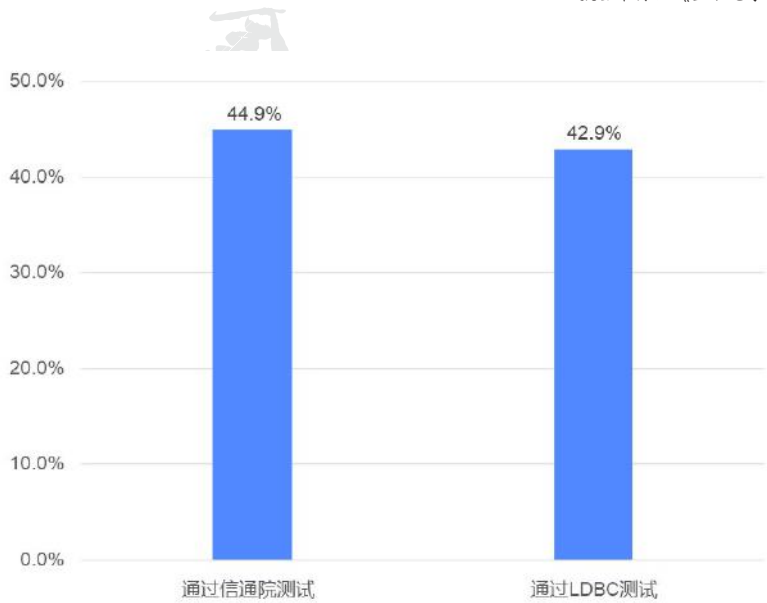
由于金融机构对于信通院测试、LDBC 测试等国内、国际图数据库测评标准了解不足，只有不足 50% 的被调研机构表示在图数据库选型过程中会考虑厂商是否通过相关测试（见图 35）。

LDBC 测试作为当前业界权威的衡量图数据库管理系统性能的测评标准，为图数据库产品测评提供了一个标准、全面的比较评判机制。LDBC 是由高校、研究所、企业联合组成的非营利组织，其中企业成员包括 Intel、Oracle、Neo4j、创邻科技等国内外知名图数据库厂商，致力于推进行业的规范标准化。SNB 是 LDBC 主导的基准测评之一，虽然采用的是社交网络数据，但模拟了真实业务场景下读写混合的任务形态，全面测试了简单交互式查询，复杂交互式查询，实时更新等



▲ 图 34 金融机构对制定相关标准的建议

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 35 金融机构对现有标准的认知情况

数据来源：《金融电子化》杂志社，2023 年 3 月

图数据库操作的性能，其中复杂交互式查询考验了图数据库对多个不同查询难点（特定通用子图结构）的查询性能，相较简单的多跳查询更贴近真实的业务查询逻辑。

此外，信通院依据《大数据图计算平台技术要求与测试方法》，开展图计算平台产品基础能力评测工作，全方位覆盖图计算平台的基本功能、兼容能力、管理能力、高可用和扩展性等能力，在 LDBC 的数据库性能测评之外，较好地完善了图计算、图分析系统的测评标准。

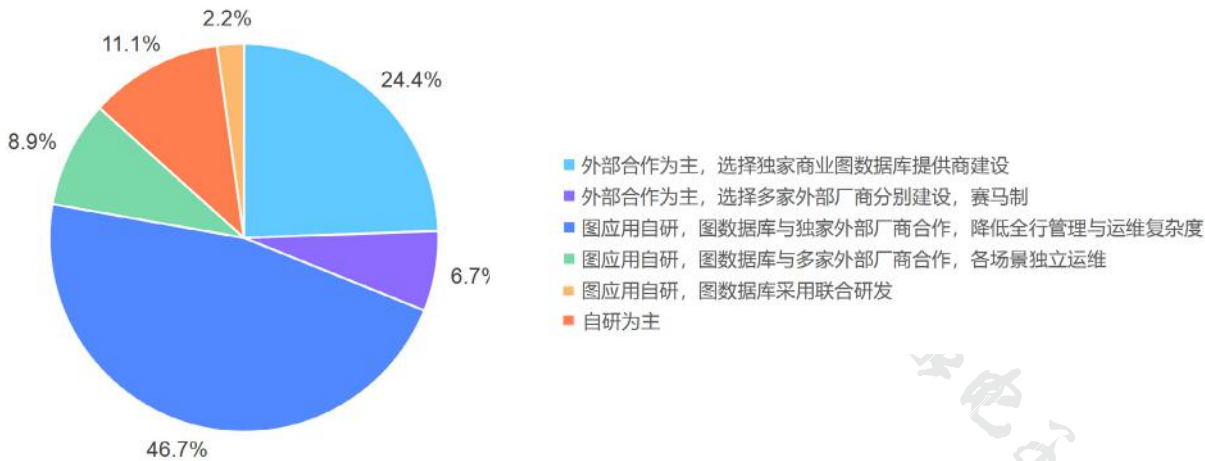
## 第四章 金融业图数据库建设与应用趋势

### 一、金融业图数据库建设整体路径

#### 1. 自研应用 + 外采图数据库成为首要建设路径

图数据库及其应用市场正处于快速发展和迭代过程，与开闭源产品相比，商业图数据库产品在性能表现、代码安全、综合运维管理成本以及成熟的服务团队等方面都具有显著优势。相对于从 0 到 1 进行图数据库自研，或是基于开源图数据库自行搭建系统，金融机构更倾向于图应用自研、图数据库与外部厂商合作的建设思路。

在 49 家计划使用图数据库的被调研机构中，45 家反馈了明确的建设思路，仅有 5 家机构表示图数据库及其上层应用以自研为主，40 家（89%）被调研机构表示会选择与图数据库厂商合作，其中，1 家表示希望以“联合研发”的方式；32 家（80%）被调研机构表示会选择单一商业图数据库供应商，以降低全行管理与运维复杂度；仅 7 家（17.5%）机构会选择与多家图数据库厂商合作，各场景独立运维（见图 36）。



▲ 图 36 金融机构图数据库建设及应用思路

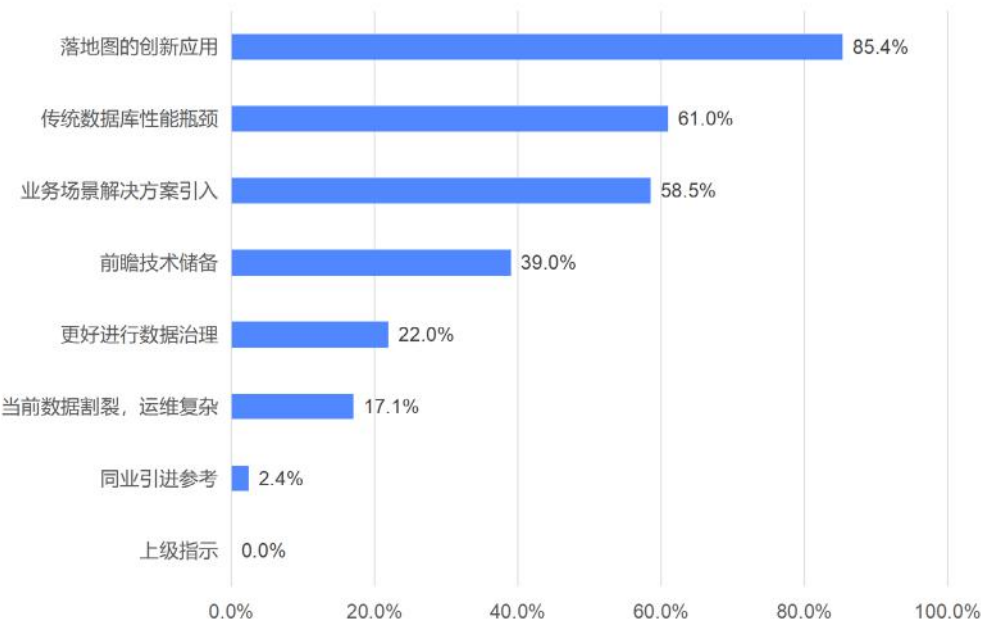
数据来源：《金融电子化》杂志社，2023 年 3 月



不同于底层图数据库，69% 的被调研机构表达了上层图应用自研的倾向。图应用非自研的 14 家机构中，78.6% 的被调研机构倾向于将图数据库及其上层应用全部外包给单一商业数据库提供商，21.4% 希望选择多家外部厂商赛马制分别建设。

2. 业务发展是驱动金融机构图数据库应用的主要动力

相比传统关系型数据库，图数据库更贴近业务应用与业务逻辑，因为它的应用意在高效挖掘多源异构数据间潜藏的内在关联、提升以客户为中心的洞察与运营能力，能够降成本、控风险、优体验、增效益。因而，业务发展成为图数据库技术应用的主要驱动（见图 37）。41 家已部署图数据库的金融机构中，85.4% 的机构表示采购决策的依据是希望落地图的创新应用，58.5% 的机构采购图数据库由业务解决方案引入。



▲ 图 37 金融机构应用图数据库原因

数据来源：《金融电子化》杂志社，2023 年 3 月

此外，随着金融机构数据量的急速增长以及关联关系挖掘需求增长，传统数据库逐渐表现出在关联关系处理上的性能劣势，61.0% 的被调研机构表示采购图数据库是由于现有场景中传统方案出现了性能瓶颈。

二、金融业图数据库建设方向

围绕金融机构规划部署图数据库的主要场景，针对 49 家计划使用图数据库的金融机构 2023



▲ 图 38 金融行业 2023 图数据库业务场景建设规划 TOP5

数据来源：《金融电子化》杂志社，2023 年 3 月

年的建设规划进行调研，数据显示，企业级图数据库平台、企业关联关系图谱、实时交易反欺诈成为前三的主流场景，占比均达到 79.6%（见图 38）。相较于 2022 金融机构图数据库建设的场景，呈现出数据跨域融通、集中管理，更关注动态实时数据分析利用的特征，可见图数据库技术应用正迈向以“大规模数据实时分析”为特征的 Graph 3.0 时代。

### 1. 打破数据孤岛，打造企业级图数据库平台

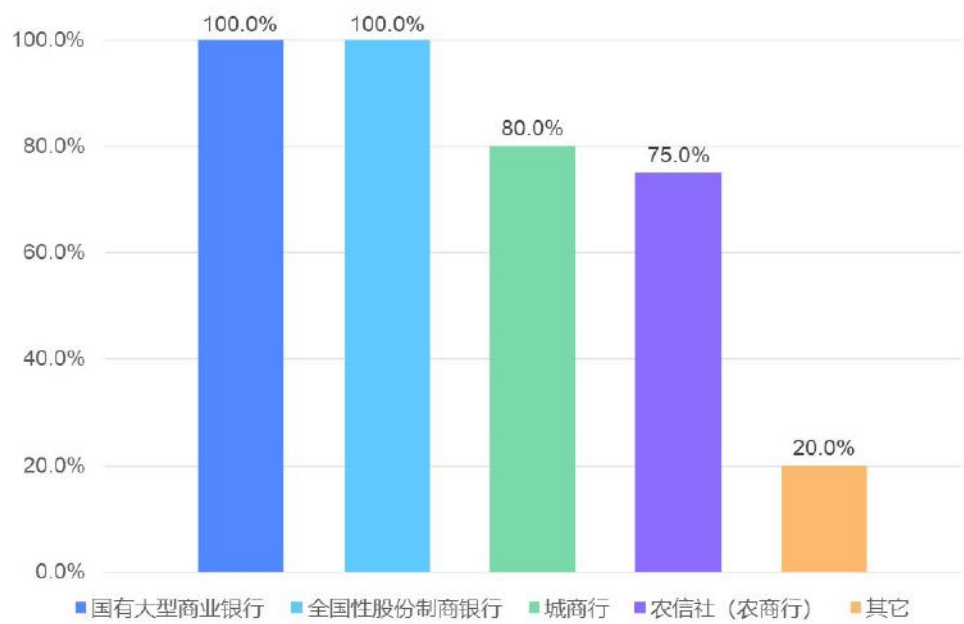
金融机构内部业务系统多元，且业务间相对独立，数据存在多、散、杂的特点，得不到很好的收集应用。业务人员需要在大量不同业务系统中寻找、清洗、整合、分析数据，技术人员需要耗费大量的人力和时间对接、运维、管理、监控不同业务系统及其后台数据库系统，数据应用及管理效率低。

79.6% 的被调研机构优先规划建设企业级图数据库平台，所有国有大行与股份制银行表示计划建设行级图平台，以图数据库技术作为技术底座，打通各部门核心数据，实现统一管理，赋能行内多数业务场景效能升级（见图 39）。

从业务侧数据应用出发，统一的行级图平台能对行内外跨部门、跨产品、跨业务线中的数据联合建模、整合融通，做到对客户数据的全局分析，支撑多元化业务。

从数据资产管理的角度，行级图平台既能够避免各业务部门重复技术选型与投入、解决跨产品间数据难寻找、难融合、全局分析耗时长、难满足实时智能决策支撑等问题。

从信息管理成本出发，行级图平台能够解决 IT 部门运维管理不同图数据库产品难度高、成本大，基于不同产品间存在的不同图查询语言、差异化的图模型、不同的图查询计算接口做应用

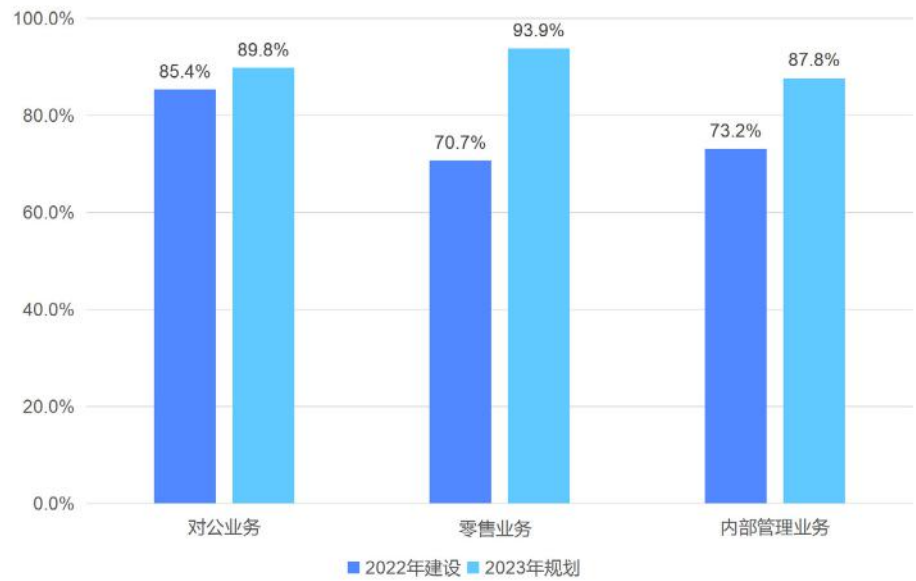


▲ 图 39 各金融机构计划建设行级图平台的占比

数据来源：《金融电子化》杂志社，2023 年 3 月

开发成本高、运维迭代难等问题。

行级图平台的建设能够高效满足金融机构对数据资产的统一管理，实现数据之间的关联关系的深度挖掘和可视化呈现，解决传统金融机构中数据利用低效、数据转换中间过程长、实时决策



▲ 图 40 金融机构图数据库业务场景应用 2022 现状与 2023 规划对比

数据来源：《金融电子化》杂志社，2023 年 3 月

难等一系列问题，全面提升金融机构的数据治理能力、营销服务能力、风险管控能力等，支撑全行“用好数，好用数”。

## 2. 图技术应用场景向多元化、实时化发展

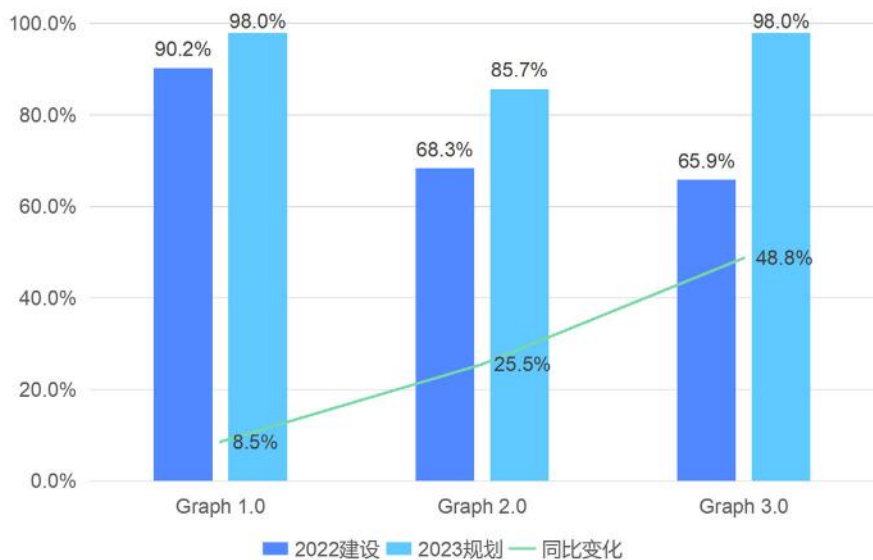
2023 年金融机构图技术的场景应用规划调研结果显示，金融机构高度关注对公业务、零售业务以及内部监管三类场景，2023 年应用规划中分别占比 89.8%、93.9%、87.8%，对比 2022 年建设现状，图数据库应用在零售场景中增长最快（见图 40）。

依据图应用特征，我们将调研的图应用场景按 Graph 1.0、Graph 2.0、Graph 3.0 进行分类（见表 2）。Graph 1.0 时代的图应用特征是数据量小、关系相对静态，如知识图谱、企业关系图谱等。Graph 2.0 时代的图应用特征是数据体量增大、但分析的时效性要求低，如反洗钱、智能推荐等。Graph3.0 时代的图应用则侧重大数据、关系动态变化、实时分析计算响应要求高的场景，如实时申请反欺诈、智能运维等。

表 2 图数据库场景

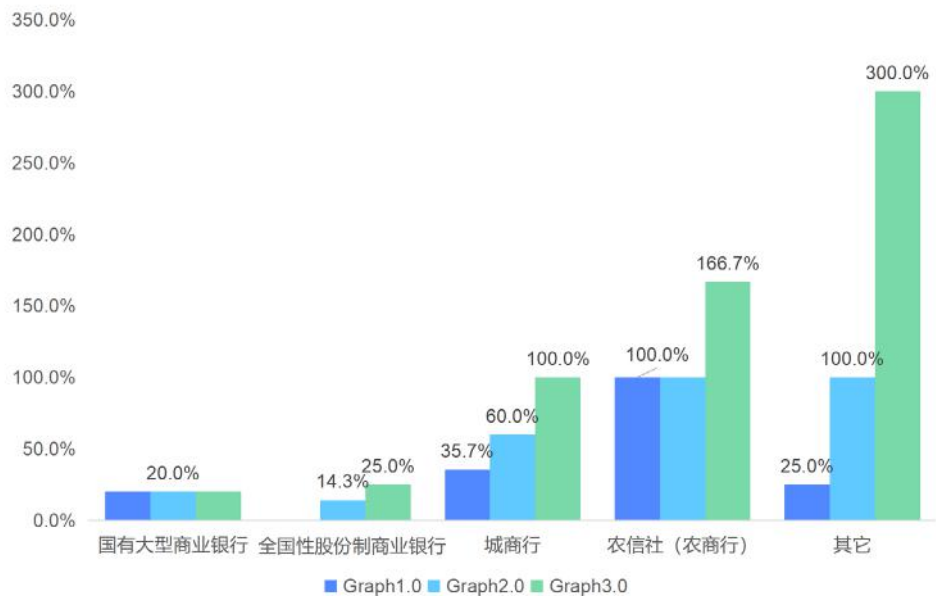
	Graph 1.0	Graph 2.0	Graph 3.0
零售	KYC：零售客户 360 图谱	信用卡反套现	信用卡申请反欺诈
	智能问答	反诈骗反赌博	实时交易反欺诈
	信用卡精准营销	智能推荐系统	
	个人业务营销	潜在 VIP 客户分析	
	失联修复		
对公	企业关联关系图谱（担保圈识别、股权穿透、集团客户管理等）		
	投研图谱		
	对公业务营销		
	对公信贷审核		
	产业链图谱		
内部监管	审计	反洗钱	网络安全
	主数据管理	外汇违规违法监管	智能运维 /CMDB
	合规内控		
	数据血缘		

针对 2022 年建设现状分析，已有 41 家机构应用图数据库，其中 37 家机构应用 Graph1.0 静态关系图谱相关场景，占比 90.2%；28 家机构应用 Graph2.0，占比 68.3%；27 家机构应用 Graph3.0，占比 65.9%。分析 2023 年的建设规划，将有 49 家机构应用图数据库，其中，有 48 家机构有规划建设 Graph 1.0 类的应用，占比 98.0%，同比增长 8.5%；42 家机构有规划建设 Graph 2.0，占比 85.7%，同比增长 25.5%；48 家机构有规划建设 Graph3.0 类的应用，占比



▲ 图 41 金融机构各类图应用场景 2022 现状与 2023 规划对比

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 42 不同类型金融机构各类图应用场景 2022 现状与 2023 规划增量对比

数据来源：《金融电子化》杂志社，2023 年 3 月

98.0%，同比增长 48.8%，金融机构的图应用在向动态构图、实时计算的场景发展（见图 41）。

分析各类图应用在不同类型的金融机构中普及率的变化（见图 42）发现，在图技术应用相

对成熟的国有大行及股份制银行中，各类图应用普及率略有提升。在图应用相对迟缓的城商行、农信社（农商行）、保险与证券机构中，各类图应用普及率同比增长巨大，尤其是 Graph 3.0 类场景的相关应用，最高达到同比 300% 增长。

### 3. 图数据库建设与应用规模增速迅猛

据 IDC 估算，2025 年全球数据总量将达到 175ZB。目前，中国金融业数据仓库已经陆续超过数百 PB 的数据量，金融机构对算力的需求也随之迈入到快速增长的阶段。

对 41 家已经使用图数据库的金融机构 2023 年图数据库的建设规划调研发现，图数据库建设在图数据规模、图的点边数量、服务器节点数量等各维度均呈显著快速提升趋势。

在图应用数据规模方面，78.0% 的金融机构规划建成 TB 量级图数据库，相较 2022 年的整

表 3 已部署图数据库的金融机构 2022 建设现状及 2023 规划

		1 节点	2-3 节点		4-10 节点	11-30 节点	30 节点
服务器节点数量	2022 建设情况	4	16		13	3	5
		(9.8%)	(39.0%)		(31.7%)	(7.3%)	(12.2%)
	2023 建设规划	2	9		15	8	6
		(4.9%)	(22.0%)		(26.6%)	(19.5%)	(14.6%)
	同比变化	-50.0%	-43.8%		15.4%	166.7%	20.0%
		1-16 核	17-32 核		33-64 核	65-128 核	128 核以上
服务器核数	2022 建设情况	8	4		7	9	13
		(19.5%)	(9.8%)		(17.1%)	(22.0%)	(31.7%)
	2023 建设规划	6	3		5	7	20
		(14.6%)	(7.3%)		(12.2%)	(17.1%)	(48.8%)
	同比变化	-25.0%	-25.0%		-28.6%	-22.2%	53.8%
		1-100GB	101-500GB	501-1024GB	1-5TB	6-10TB	11TB 以上
存储规模	2022 建设情况	7	3	7	12	5	7
		(17.1%)	(7.3%)	(17.1%)	(29.3%)	(12.2%)	(17.1%)
	2023 建设规划	3	5	1	16	3	13
		(7.3%)	(12.2%)	(2.4%)	(39.0%)	(7.3%)	(31.7%)
	同比变化	-57.1%	66.7%	-85.7%	33.3%	-40.0%	85.7%
		0-1 百万	1 百万 -1 千万	1 千万 -1 亿	1 亿 -10 亿	10 亿 -100 亿	100 亿以上
图谱点边数量规模	2022 建设情况	3	5	8	15	6	4
		(7.3%)	(12.2%)	(19.5%)	(36.6%)	(14.6%)	(9.8%)
	2023 建设规划	2	4	2	16	10	7
		(4.9%)	(9.8%)	(4.9%)	(39.0%)	(24.4%)	(17.1%)
	同比变化	-33.3	-20.0%	-75.0%	6.7%	66.7%	75.0%

体建设规模（58.5% 机构达到 TB 量级），同比增长 33.3%，建成 TB 量级图数据库的机构中，31.7% 的机构图数据规模将达 10TB 以上，同比增长 85.7%。

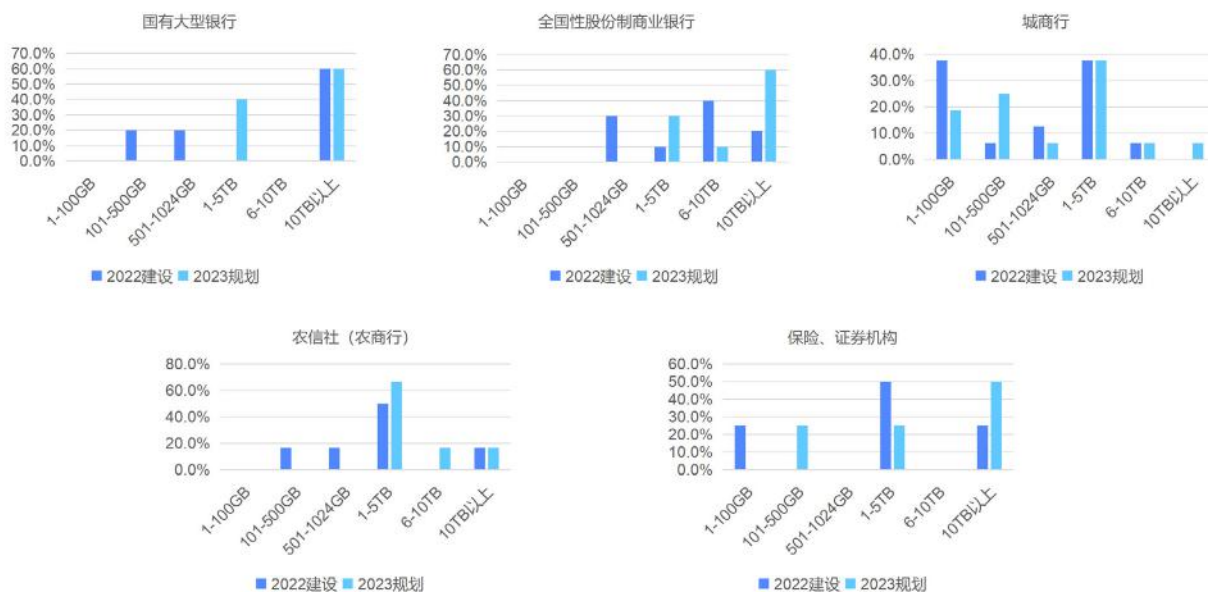
在图点边规模方面，41.5% 机构规划建成 10 亿规模以上大图，同比增长 70.0%，其中，17.1% 达 100 亿以上规模，同比增长 75.0%。

在使用服务器节点规模方面，主流建设规模从 2-3 节点升级为 4-10 节点，仅 26.8% 的机构建设规模小于 4 节点，同比降低 45.0%。34.1% 的机构计划建设超 10 节点的图数据库集群，同比增长 75.0%。

在使用的服务器核数上面，48.8% 的机构预计采购 128 核以上大算力服务器，同比增长 53.8%。可见在金融机构图应用数据量激增的同时，对大数据的关联查询、分析性能也提出了更高需求，金融机构数字化转型、高效挖掘数据资产价值的原生需求强烈（见表 3）。

分析不同类别的金融机构规划的图应用数据规模（见图 43）可见，已使用图数据库的国有大行、股份制银行、农信社（农商行）2023 年已经全面达到 TB 级别图数据应用规模，60% 的股份制银行已达到 10TB 以上数据规模，同比增长 200%。在城商行与保险、证券机构中，整体数据量呈上涨趋势，但机构间差异依然较大，从小于 100G 到 10TB 以上的机构都有分布。

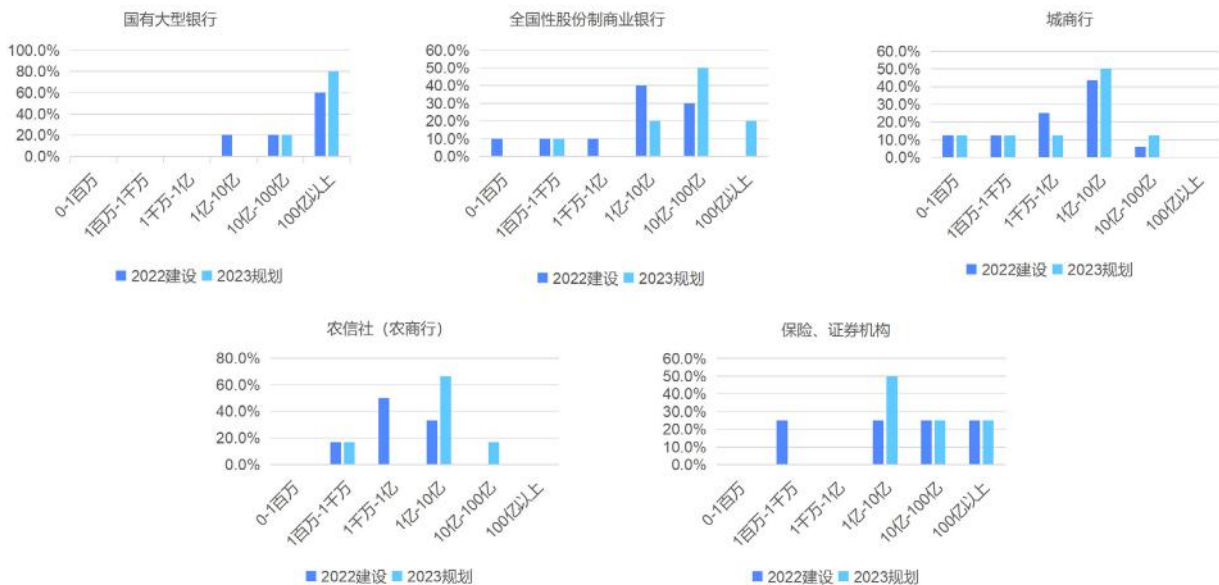
分析不同类别的金融机构计划投入的图数据库集群规模（见图 44）可见，除股份制银行的集群规模变化不大外，其他各类已使用图数据库的金融机构普遍计划扩容。城商行普遍从 2-3 节



▲ 图 43 不同类型金融机构 2022-2023 年图应用数据规模建设对比

数据来源：《金融电子化》杂志社，2023 年 3 月





▲ 图 44 不同类型金融机构 2022-2023 年图数据库集群规模建设对比

数据来源：《金融电子化》杂志社，2023 年 3 月

点扩容至 4-10 节点，60% 的国有大行达到 30 节点以上超大集群规模，同比增长 50%。保险与证券机构最快，从 2022 年 100% 的机构集群数在 4-10 节点间，变成 2023 年 75% 的机构提升到 11-30 节点的集群。

### 三、金融业期待与展望

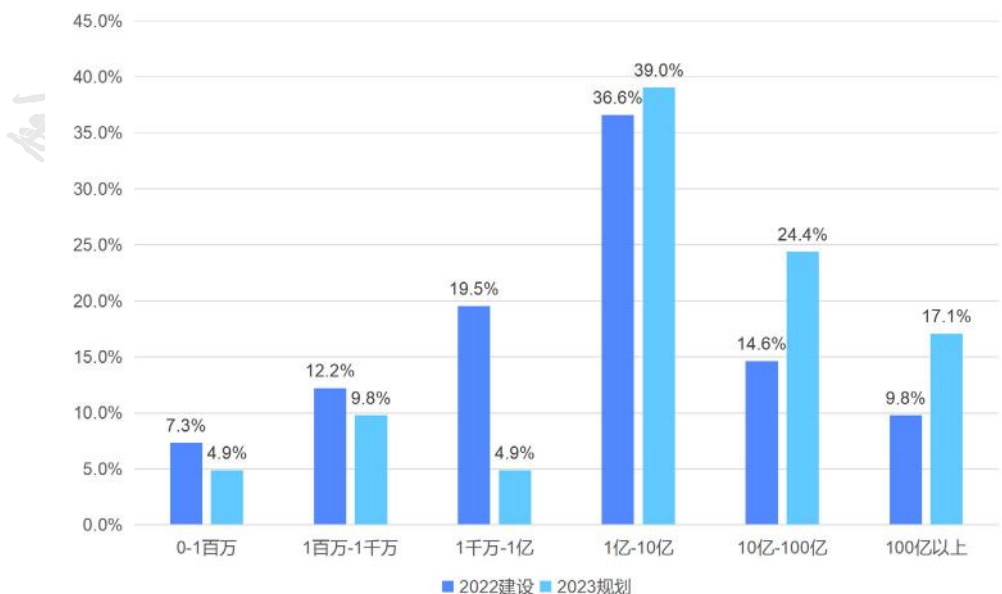
图技术作为理解世界的新方式，正凭借其复杂关联关系的强大刻画能力赢得更多金融机构的关注。调研发现，金融机构期待图数据库技术朝着以下方向迭代发展：

#### 1. 大规模图数据实时分析需求增加、推动图技术向 Graph 3.0 时代升级

随着数字经济的高速发展，数字金融的大规模普及，金融机构的数字化转型加速，图数据库的应用逐渐从静态的关系数据扩展到动态交易、申请、行为、事件数据，同时存储、分析和计算的数据量也呈爆炸式增长。

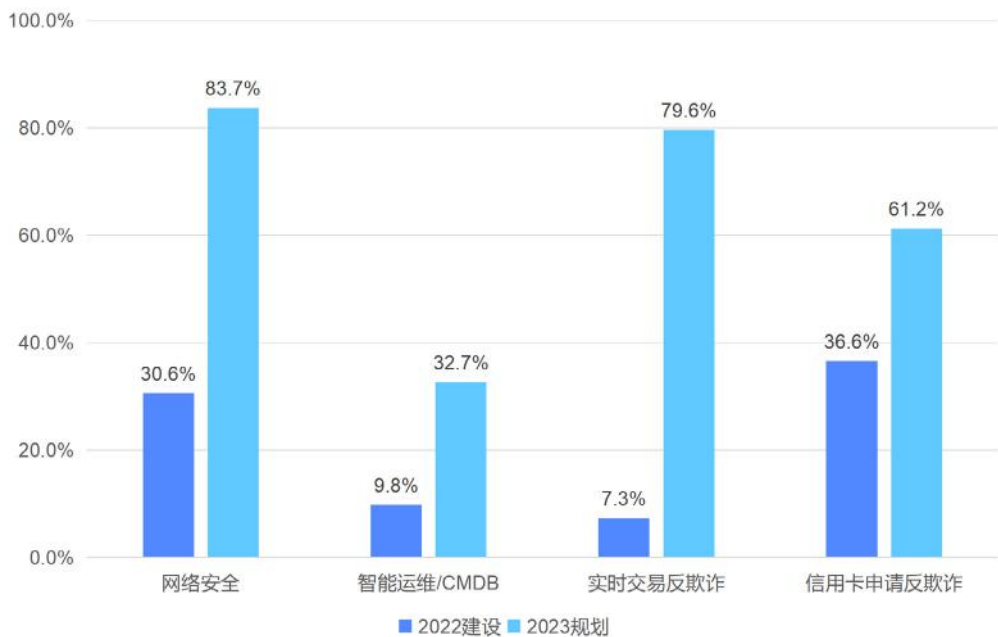
一方面，金融机构内部积累的数据量呈井喷式增长，图规模亦快速增长。对已经使用图数据库的 41 家金融机构调研，结果显示，2023 年 80% 以上的被调研机构图规模将超 1 亿点边，近 20% 的被调研机构的图规模将超过 100 亿点边，图应用深度与广度在迅速增长（见图 45）。

另一方面，金融机构的业务日趋争分夺秒，无论是反洗钱、反欺诈，还是交叉营销和消费升



▲ 图 45 金融机构 2022-2023 图规模发展

数据来源：《金融电子化》杂志社，2023 年 3 月



▲ 图 46 Graph 3.0 典型场景图数据库应用趋势

数据来源：《金融电子化》杂志社，2023 年 3 月

级，都期待突破 T+1 的桎梏，实现实时响应。以 Graph3.0 的典型场景智能运维为例，金融机构需要对 IT 基础设施和应用程序的自动化监控和实时管理，以便及时发现并应对系统故障，确保业务系统和应用程序的高可用性、可靠性和安全性，从而保障客户的服务体验。调研结果显示（见图 46），分别有 83.7%、32.7%、79.6%、61.2% 的金融机构计划于 2023 年将图数据库应用于网

络安全、实时交易反欺诈、信用卡申请反欺诈、智能运维等高实时性场景。其中，以实时交易反欺诈的应用增长最快，同比增长 72.3%。网络安全作为全新的图应用场景进入金融机构视野。

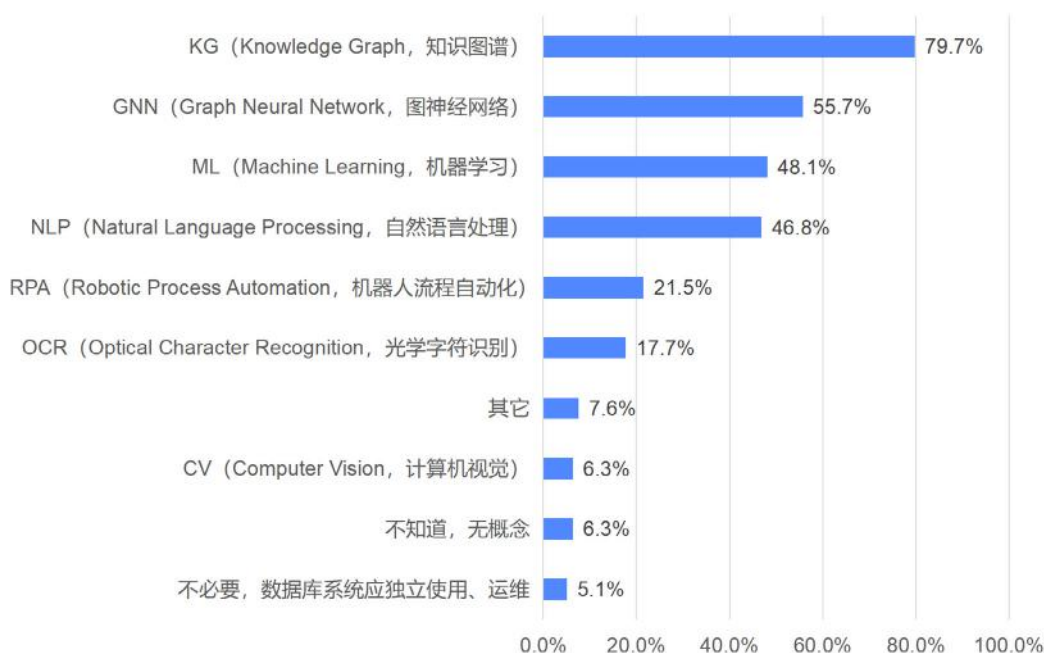
在金融图应用逐步迈向 Graph 3.0 时代的大背景下，机构对更大规模数据的实时分析需求必然推动商用图数据库向着承载数据更多、查询计算速度更快、算法支持更丰富等维度演进，可预见具备查询快、运算快、高扩展的第三代原生分布式图数据库将成为金融机构的优先选择。

## 2. 与 ML、DL 等人工智能技术结合，实现可解释的 AI

随着金融业务日益复杂，金融机构面临着提高业务决策的准确性和可解释性挑战，传统机器学习模型缺乏决策依据的透明度和解释力。大量实际应用场景中的数据从非欧式空间生成，传统深度学习方案无法高效处理。

为解决以上问题，图数据库与人工智能相结合成为具有潜力的解决方案。图数据库能够提供高效和灵活的方法存储和管理图数据，并提供丰富的查询和分析功能，以便于人工智能应用程序从中获取有用的信息和知识用于训练模型、推断过程和预测结果，提升系统的实时性和准确性，有效避免“黑天鹅”和“灰犀牛”事件的发生。

近年来，图 + 机器学习、图 + 深度学习在人工智能领域引起广泛的研究兴趣。近 90% 的被调研机构期待图数据库与 AI 结合，以 KG 知识图谱（79.7%）、GNN 图神经网络（55.7%）、ML 机器学习（48.1%）、NLP 自然语言处理（46.8%）等技术受关注度最高（见图 47）。



▲ 图 47 金融机构对图数据库与人工智能结合的看法及期

数据来源：《金融电子化》杂志社，2023 年 3 月

### 图数据库技术 + 知识图谱

据艾瑞统计测算，2021 年 AI+ 金融领域知识图谱核心产品市场规模达 21 亿元，预计 2026 年核心产品市场规模达到 68 亿元，2021-2026 年，年复合增长率高达 26.6%。

图数据库和知识图谱是两个相关但不同的概念：图数据库是一种存储图数据的技术，知识图谱则是用实体、关系、属性表示和组织信息的方法。知识图谱能够从应用层将客户行为、关系等建立关联分析，结合丰富的可视化能直观地呈现出数据背后所隐含的推理逻辑、提供决策依据。图数据库与知识图谱的结合可以支持高效的信息检索和复杂的数据分析，如基于知识图谱的推荐、利用知识图谱增强的搜索引擎等。两者的结合，让金融机构能够从技术层到应用层形成完整的解决方案闭环，更快实现业务价值落地。

### 图数据库技术 + 图神经网络

图神经网络是指使用神经网络来学习图结构数据，提取和发掘图结构数据中的特征和模式，满足聚类、分类、预测、分割等图学习任务需求的算法总称。它将人类已有的认知以图的形式建模并与计算机建立的神经网络模型相结合，有助于解决传统机器学习算法存在的可解释性低下、缺乏非欧空间数据结构处理能力等问题，已成功应用于推荐系统、计算机视觉等领域。

随着图数据应用越来越普遍，其蕴含的信息越来越丰富，图神经网络的应用场景也会越来越广泛。图神经网络和图数据库的结合，能够将深度学习的预测能力应用于丰富的图数据结构上，帮助金融机构对数据进行全量推理，训练预测模型，从图中挖掘出全新的可解释的逻辑，获取正向业务知识反馈。以金融风控场景为例，基于图神经网络和图数据库技术的结合，金融机构能够清晰、实时地了解各个数据间的业务关联和风险传导路径和风险识别规则，从而精准定位、识别、判定风险，推动金融风控向更精准、可解释智能的维度发展。

### 图数据库技术 + 机器学习

目前，部分图平台产品已经实现了对机器学习相关功能的支持与扩展，除常规图算法之外，也开始支持与第三方机器学习平台的对接，通过将图挖掘与机器学习算法相结合，提高模型的准确率与可解释性。但图机器学习领域尚在发展早期，仍存在内存、硬件等多方面瓶颈，未来在大规模图挖掘的工程化方面仍有很长的路要走。

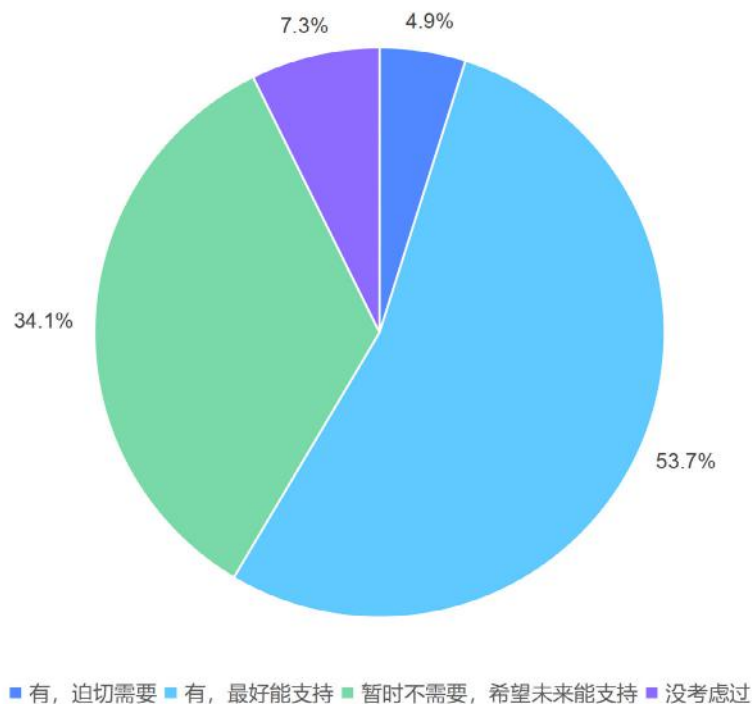
### 3. 关注时序数据分析，实现业务动态化分析

数字化进程的加速推进，要求金融机构对数据的处理能力从事后的“跟随分析”变为实时的“伴随分析”，从响应式风险管控变成干预式风险防控，从标准金融产品的推送变成以客户为中心的定制化金融产品设计。这必然要求金融机构实时收集、处理、分析大量的用户行为、事件数据。

时序数据是指一段时间内采集的，按时间顺序排列的数据，最大特点是数据具备时间属性。

金融场景中，存在大量的时序数据，交易转账、产品购买、客户访问、价格波动等，都具备很强的时间特性。对这些数据的高效监控、分析能够支撑金融机构更实时、准确的业务决策、提升风险管理效率，甚至发现新的市场机会，创新产品和服务以满足客户需求。

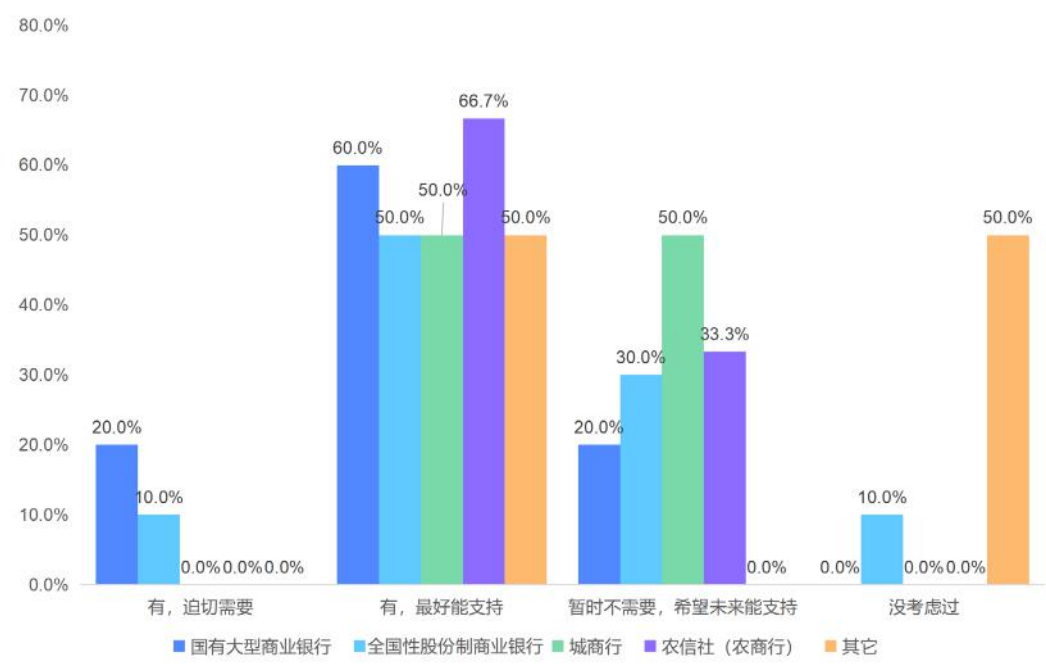
利用时序数据构建金融时序信息图谱（交易图谱、申请图谱、运维图谱等），通过将历史数据按时间段切分，实现对数据关联关系的动态变化的分析，得出基于时间维度的分析结论，立体、直观发现风控、营销、智能运维等业务实践及 IT 管理过程中的动态关系及属性特征变化，对未知的可能性进行预估判定，真正做到对分析对象的形成性、动态认知。被调研的 41 家金融机构中，53.7% 表达了对图数据库支持时序数据分析的明确需求，另外 34.1% 哪怕暂时不需要也希望未来能支持。这也进一步体现了图应用正逐步迈向基于行为、事件的实时分析的 Graph 3.0 时代（见图 48）。



▲ 图 48 金融机构对图数据库时序分析功能看法及期待

数据来源：《金融电子化》杂志社，2023 年 3 月

不同类型机构对时序分析的态度也体现出一定的不同。国有大行对于时序分析的需求最为强烈，80% 的国有大行均期望建设时序功能（见图 49）。



▲ 图 49 不同类型金融机构对图数据库时序分析功能看法及期待对比

数据来源：《金融电子化》杂志社，2023 年 3 月



## 第五章 图数据库未来发展建议

### 一、底层技术自主可控成为关键竞争力

我国的金融业务体量和场景的复杂度，均处于世界前列。随着新兴技术更深度地参与到金融交易、风险决策等活动中，金融风险的复杂性、传染性、交叉性进一步加剧。在当前多重因素影响下，传统、固化的底层软硬件设施已经无法满足最新的数据安全需求。

对于金融行业而言，基础设施决定了业务场景的安全可控、高效智能。因此，金融机构积极响应国家自主创新的号召，加速提升各类软硬件的自主可控能力，打造底层架构，提升底层设施的安全能力，为形成高度稳定级的上层应用打好稳固的基础。

未来，金融机构在进行图数据库选型时会更加关注图数据库的自主可控能力，并且关注其和其他软硬件的适配能力，实现核心业务系统的成功攻关，打造分布式核心平稳、安全的运营能力，解决关键能力瓶颈。

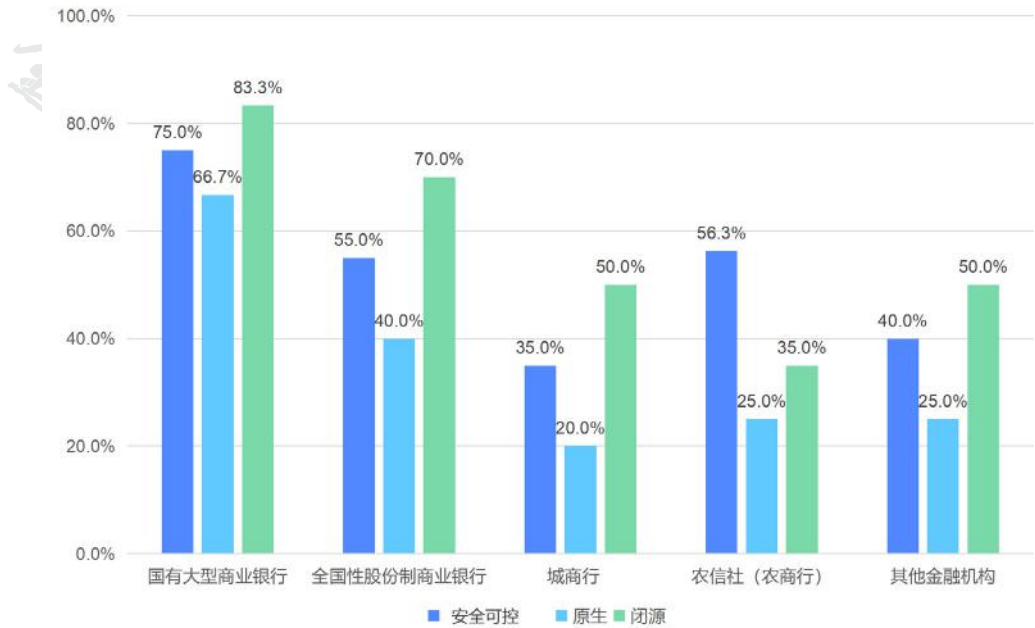
当前，国有大型商业银行开始应用图技术赋能行内业务，拥有更成熟的数据库应用，更注重图数据库的安全性与技术自主性（见图 50）：75.0% 的机构要求必须为自主可控的图数据库，83.3% 要求原生，66.7% 要求闭源，对图数据库厂商安全自主可控提出了更高的要求。

### 二、人才是打通图数据库金融行业应用的关键

图技术应用市场的繁荣发展离不开人才储备。图数据库领域人才的稀缺已经成为金融机构面临的共性问题。78.5% 的被调研机构表示缺乏相关专业人才更好地使用图数据库，业务部门对相关人才的缺乏度达到 44.3%（见图 31）。

对业务部门而言，拥有成熟专业的咨询团队，能够提供包括方案评估、设计、交付、迁移、测试工作过程等在内的规划服务，是图数据库厂商产品高效落地的关键竞争力。图数据库厂商在不断优化产品能力，提升产品性能外，需要重视图技术应用综合人才的招聘与培养，加强与下游图解决方案商的生态合作，降低金融机构应用、发掘图数据库价值的门槛和成本，打通图数据库应用“最后一公里”。





▲ 图 50 不同类型金融机构对安全可控的要求

数据来源：《金融电子化》杂志社，2023 年 3 月

被调研机构表示期待有更多机会获得图数据库厂商的技术指导，其中，获得图数据库场景应用支持需求占比 63.3%（见图 51）。

因此，图数据库行业中的各厂商，可以通过开设图数据库基础理论知识和实践案例分析的培训课程、打造图数据库实战项目实践平台以及定期举办图数据库技术交流活动等宣传动作，加强图数据库业务应用和技术开发人才的培养力度；并且面向金融机构打造系统化的图数据库知识培训体系，利用系列培训课程和测评考试等方法，培养更多懂得金融业务的实战型、咨询型人才，赋能金融机构更方便、更快捷地使用图数据库，发挥图技术价值。

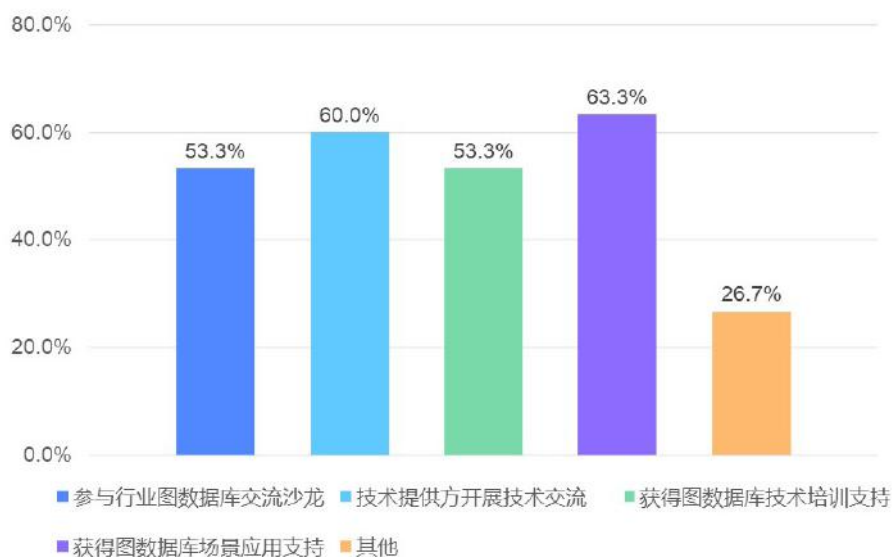
### 三、存储、计算、分析一体化的图平台或成主流

随着金融机构图应用规模的逐步增大、图应用场景逐步迈向 Graph 3.0 阶段，需求从离线 / 准实时计算到实时计算，金融机构对图数据库厂商的动态大规模图数据处理能力提出了更高的要求，推动图数据库产品迈向集图存储、计算、分析一体的平台化趋势。

在传统解决方案采用的“图数据库 + 图计算引擎”独立系统架构下，对计算系统而言，存储系统内部通讯机制是黑盒，无法做执行过程优化，不仅查询、计算效率低，更重要的是数据从存储系统传输到计算系统、计算结果持久化回存储系统的时间较长，无法完成对实时写入的大数据进行复杂图查询与计算的任务。

相较之下，一体化图平台可以融合图数据库的 OLTP 能力与图计算系统的 OLAP 能力，通过一致性协议集成为新一代 HTAP 图平台，而 HTAP 图平台在原生分布式图架构下将会实现更好的性能、扩展性和可靠性。

以 Graph 3.0 的国内代表产品 Galaxybase 图平台为例，它采用分布式原生图存储的架构，能够根据图数据特征或动态或静态地进行分片管理，优化数据跨节点分布情况；并且依据底层数据存储的情况，高效实现各种分布式计算任务的调度和分发策略，降低跨节点数据网络传输与通信成本，实现存储计算协同，提升整体图查询、计算的性能。它在满足水平扩展的前提下，提升了查询与分析效率，能更好地支持金融机构海量行为、事件数据构成的超大规模业务图谱的高效查询分析和稳定运行。



▲ 图 51 金融机构对图数据库宣传推广态度

数据来源：《金融电子化》杂志社，2023 年 3 月

## 第六章 金融业图数据库实践案例

### 农业银行图数据库建设历程

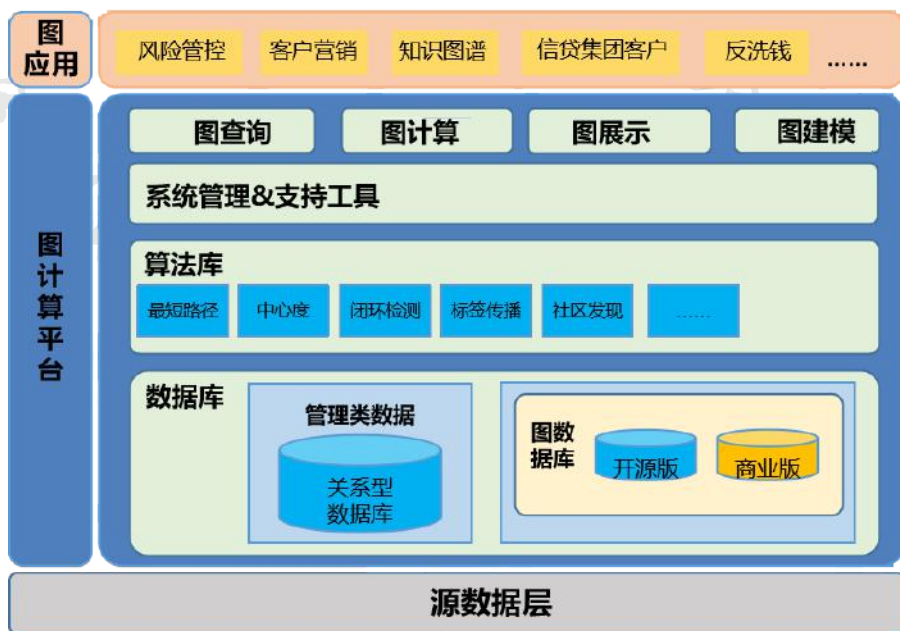
中国农业银行研发中心 罗晓峰 李博

现如今，数据分析越来越复杂、模式更改越来越频繁，银行业内多重关联分析、知识图谱建设、数据价值深度挖掘等需求越来越旺盛。传统关系型数据库已无法完全满足关联数据分析处理的挑战。图数据库通过图结构组织数据，克服了其他数据库难以支撑的深度关联数据分析痛点，为构建银行知识图谱，实现深度业务知识和价值挖掘提供重要科技保障和技术指引。为此，农业银行图数据库技术引入工作于 2018 年正式启动，经过几年来的研发实践，图数据库技术在农行信贷监测、反洗钱、风险控制、系统运维等领域均落地了较为成熟的应用场景，充分利用图数据库的数据挖掘能力，为业务创新提供基础支撑。

#### 初步探索——建设图计算平台

如图所示，作为行内图数据库技术应用的前沿，农行图计算平台最初设计并建立了图查询、图计算、图展示、图建模四大服务模块，集图构建、图存储、图分析建模、图管理、图展示于一体，提供实时和批量图计算服务、可视化及其他辅助工具，支撑各类图计算应用。建设伊始，同业之中尚无先例可循，对于如何建设一个能够为全行提供图计算服务的综合服务平台，图计算平台总体设计成为了首先面临的关键问题。

从发展策略上，图计算平台底层图数据库使用坚持商业开源双线并举。自 2019 年开始，农行基于开源图数据库，探索搭建图计算平台，支撑全行应用系统深度关联数据分析处理，为农行提供基础图存储、图计算公共服务。2020 年，图计算平台开始商业版图数据库引入工作，采用商



▲ 图 图计算平台技术架构

业版图数据库来强化图数据库核心，健壮图数据库服务能力。这样对于不同需求的应用系统来说，可以通过图计算平台来选择适合自身现状的图数据库，满足差异化应用场景建设。

从能力建设上，图计算平台采用前后端分离架构，对外提供联机接口、辅助开发建模工具，支持各系统实时查询、更新图数据，实现对外服务能力；采用“关系型数据库+图数据库”混合存储架构，管理型数据使用关系型数据库进行存储，应用分析类数据使用图数据库进行存储，同时发挥两个数据库在不同功能上的优势，自行设计主备机高可用架构，保障系统高可用能力；利用图数据标签建立了基于子图的用户权限体系，实现对各上层应用数据权限管控和数据隔离，实现一定安全管控能力；采用分层设计，制定了图应用数据交互接口标准，屏蔽底层图数据库语言差异，从平台层面规避图查询语言尚未统一标准的问题，提升应用系统适配能力。

图计算平台的初期探索建设，实现了农行在图数据库技术应用方面从0到1的过程，虽然图计算平台仍有许多方面需继续完善，但还是为行内进一步深化应用图数据库技术打下了基础。

### 持续提升——图计算平台能力完善

为更加深入应用图数据库技术，使之能够在行内大规模推广，农行通过制定相关技术规范，增强运行保障能力，丰富场景支撑，对图计算平台和商业版图数据库功能持续进行优化提升。

在图计算平台不断完善的过程中，农行联合厂商制订了图数据库在行内的技术规范，只有这样，图数据库才能够具备在行内大规模推广使用的条件。例如，在安装部署规范中，不同等级系

统在使用图数据库时的部署架构是不同的,规范必须明确不同等级系统对应的图数据库集群架构,否则系统无法知道应该使用什么样的架构;又例如,安装部署规范还需要明确图数据库的容量模型,以便应用可以根据自身情况申请图数据库资源。此外,还有健康检查、系统监控、应急处理等一系列规范,均需要明确才能保证图数据库在行内具备推广的基础能力。除了行内的规范制定,图数据库在大型企业的应用标准仍然是较为空白的一个领域,农行根据自身实践经验,正在牵头制定《图数据库金融应用技术要求》,旨在补全图数据库技术在业内应用标准的空白。

随着应用图数据库技术的系统越来越多,图数据库的运行保障不再是对简单的单一实例的维护。图计算平台为了应对多集群、多用户、多场景的运维挑战,新增了图项目管理、集群管理、集群监控、资源管理等一系列运维保障功能,提出基于图项目的租户模式,调整图计算平台服务架构,解耦平台应用边界,释放应用系统场景建设和技术研发潜力。通过进一步明确资源隔离模式,图计算平台实现了对不同应用的精细化运维管理。

场景支撑方面,通过对底层图数据库的不断升级优化,采用原生分布式并行图存储,支持千万级邻居秒级查询;利用分布式架构,支持动态在线扩容,高效支持大规模数据量存储,图计算平台提供服务以来,数据规模已从开始试点时的百万级点边数据扩展到十亿级点、百亿级边的规模。

## 推广应用——多领域场景落地

在图计算平台的支持下,农行图数据库应用稳步前行、逐渐深入。信贷风控、智能营销、内控审计、系统运维、知识图谱等领域均已落地相关应用,图数据库应用已取得阶段性成果。

**1. 集团客户风险监测。**相比传统银行授信业务,集团客户授信需要对集团公司及各个子公司间关联关系进行深入分析,并在单一企业风险分析维度的基础上从集团整体的角度进一步分析集团整体的风险。图数据库技术优异的关联数据深度分析能力能够帮助业务人员在这一传统业务中获得新的技术支撑。“集团客户风险监测”作为行内精准识别十大应用之一,基于图数据库技术建立了集团客户关联关系分析和风险分析两类模型,实现对股东关系的10层股权穿透,挖掘隐性集团、担保圈、资金圈等隐性关系,两模型总体准确率分别达到82.94%、81.99%,精准定位隐性集团客户与风险集团,为各级行防范与化解大额风险提供精确线索。实现了农行信贷风控从“专家经验”到“智能识别”的转变,极大提升了农行的信贷风险预警和防控水平。

**2. 智能客户精准营销。**基于传统关系型数据库的客户、产品的深度关系难以挖掘,很大程度上制约了对客户的精准营销。农行利用图数据库技术,将存量客户、目标客户以及资金链、关键人等关联关系的分析变为现实。根据分析结果,构建客户画像,从中挖掘行内外高价值客户,为

业务人员提供潜在目标客户名单，助力营销能力提升，进一步解决了客户在哪、存款在哪、业务在哪的问题。提升优质重点客户的信息获取能力，有利于及时、有效地开展精准营销，提高行外客户营销转化工作效率，增强行内客户黏性，实现客户资金在行内封闭运行，增强了农行综合收益。

**3. 内部审计交易图谱检索。**在银行业，内部审计监督工作是保障和促进资产保值增值的重要一环，审计人员约一半以上的审计查证模型是基于交易流水数据展开的，然而通过 SQL 规则加工方式来描述账户间的资金交易情况，面对复杂资金往来关系时，存在加工逻辑复杂、加工时间漫长、开发效率低下等痛点。农行利用图数据库技术，在内部审计管理监测系统搭建了基于图的交易图谱查证场景，很大程度上解决了当前资金交易关系查证难的问题。

**4. 应用系统关联关系分析。**对于农行这样一个巨大规模的企业，拥有成百上千套 IT 系统需要运维管理，这些系统之间往往不是孤立存在的，而是有着错综复杂的依赖关系。如果一个模块出现故障，则可能影响一个或多个下游系统的运行。图计算平台利用图数据库，清晰地描述这些关系，当一个系统模块发生故障时，能够快速分析出可能受到影响的下游系统，应用系统关联关系分析为农行提供了一种高效的系统故障影响分析方法，极大地提升了企业的运维水平。

**5. 知识图谱。**基于图数据库，农行构建了知识图谱引擎，旨在持续积累全行领域级图谱资产，覆盖自然、信贷、金融资产等多个维度，累计建设全行个人、法人、产品、文章知识等数十类关系图谱。通过知识图谱构建、挖掘和积累，结合领域专家知识，提升已有模型准确性和发掘新的业务规则。例如，依托知识图谱引擎进行反洗钱可疑客户识别，应用系统基于 3 个月交易流水信息构建了包含 1.6 亿实体和 3.7 亿关系的图谱，最终筛选出约 2000 多个可疑客户，准确率超过 94%。



## 交通银行全栈式知识图谱平台及应用建设

交通银行金融科技部

根据行业发展趋势，当前金融企业在对客户、产品及行业进行分析时，往往不能仅靠单一的规则及因素，需结合多方面、多层级、多维度的关联关系，进行综合分析和深度挖掘。近年来，知识图谱技术发展迅速，其主要特点为：将客户等各种信息建立立体的关系图（包括集团关系、股权关系、投资人关系、交易关系、担保关系、诉讼关系、事件关系等），再以图分析和图计算为基础对信息进行挖掘，发挥数据深层价值。

交通银行于 2019 年引入业界成熟的商用图数据库，并在此基础上自研全栈式知识图谱平台，覆盖数据获取、图谱构建、图谱管理、图谱应用等关键环节，具备全栈式图谱生产与应用服务能力。该平台建成后，有效地支撑风控、营销等领域的多个业务场景，并取得一定的业务成效。

### 全栈式知识图谱建设方案

随着知识图谱技术在交通银行业务应用的拓展，需要重点提升该技术的应用效率。通过实践，影响图谱技术应用规模化的因素主要有以下三点：

一是数据复用难。过去，一般采用从特定需求出发构建图谱及其应用的建设模式，这种“烟囱”模式无疑能带来最大程度的解耦，但构建出的图谱异质且分散，难以集成与复用。

二是应用门槛高。图谱分析需要掌握图查询语言、图计算框架以及各类图算法，对非技术人员不友好，影响图谱应用的推广。

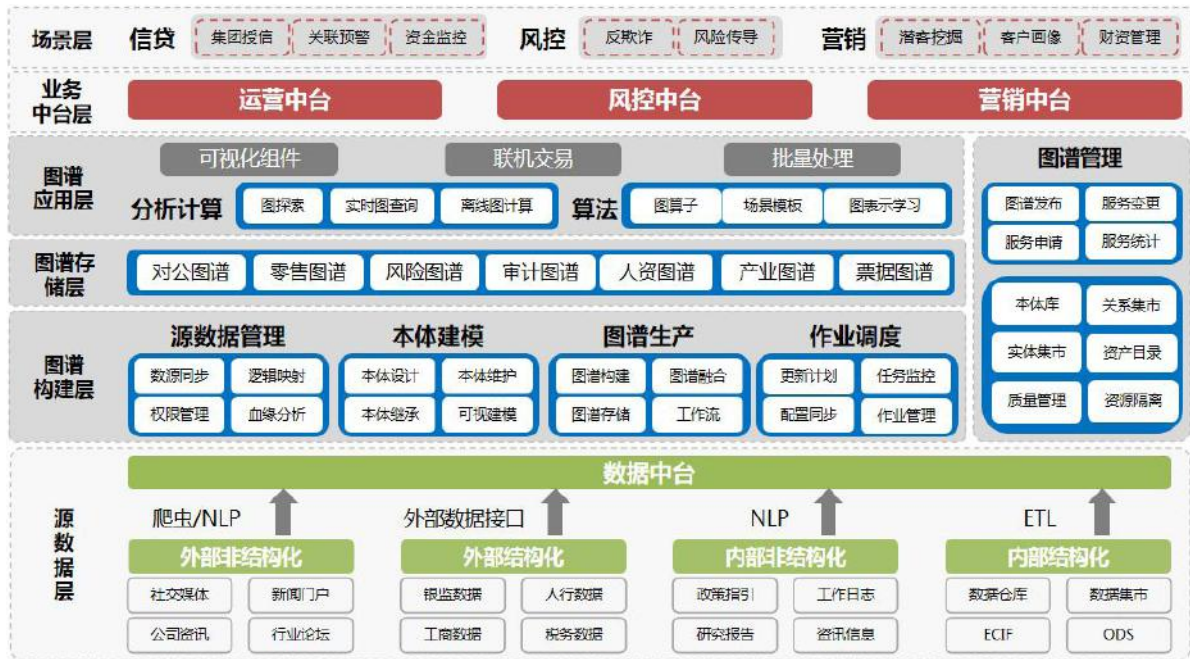
三是开发环节多。图谱建设涉及到数据收集、数据清洗、数据处理、图谱构建、图谱应用等流程，自动化程度低，导致图谱生产效率低下，无法快速响应银行日益增长的业务需求。

为解决这些痛点问题，交通银行基于微服务架构和自研的微服务治理平台建设全栈式知识图谱平台，为业务提供一套集图谱构建、图谱管理、图谱应用于一体的敏捷的全流程闭环生态平台。

在图谱构建层，交通银行自主研发图谱生产工作流引擎，提供基于可视化界面的图谱生产工作流服务，支持在 DAG 画布中通过拖拉拽方式组合不同的图谱生产算子，提高开发效率。

在图谱管理层，通过本体库、实体集市、关系集市、资产目录、图谱发布等管理功能，为图





▲ 图 交通银行全栈式知识图谱平台逻辑架构图

谱构建提供规范化的指引，解决图数据融合困难的问题。

在图谱应用层，通过运用 Cytoscape 图谱可视化组件，自主研发可视化图查询引擎，为用户提供自定义图形化查询工具，并预置图查询算子，降低平台使用门槛，提高用户友好度。

基于全栈式知识图谱平台形成“图谱构建 - 图谱管理 - 图谱应用”的完整生态闭环，实现知识图谱从简单使用到生态打造的跃迁，有效缩短图谱构建周期，对于及时响应市场需求，提升金融服务质效有着重要意义。

## 知识图谱应用建设

交通银行借助知识图谱技术，深度挖掘关联关系，并与机器学习结合，在风控、营销等领域先后为风险监测、反洗钱、反欺诈、贷后资金监管、联动拓客、产业链拓客等场景建设一批智慧化应用，业务成效显著。

在风控领域，交通银行致力于建立“全集团、全覆盖、全流程”的智能化风险监测体系，充分考虑企业风险事件的关联性、传播性等特点，构建出以关联关系为基础的风险预警模型，将风险事件的预警从单体识别变成群体识别，切实提升风险排查精度、风险预警敏感度和应急响应速度，有效防范系统性风险、区域性风险和重大个案风险。一是识别集团客户管理漏出，掌握集团客户在交通银行的全口径融资情况。目前集团管理漏出预警已经嵌入交通银行授信发起流程，并在风

险监测中实时提示，有效改善分行集团授信管理中应建未建和应纳未纳问题。二是探查企业隐性关联关系，提高客户洞察能力。针对交通银行全集团授信业务有贷户组合进行预测，挖掘出确实存在隐性关联的客户，解决股权信息未公开企业之间隐性控制关系难以发现的问题，为客户准入、集团授信管理、贷后风险预警、不良处置等决策提供有力参考。三是完善担保网络风险监测，识别和防范担保链圈造成的区域性风险。完成对全行授信客户担保联通体的风险排查，将之前未纳入黑灰名单管理的潜在风险余额纳入名单制管理，有效防范风险传染。

在营销领域，交通银行基于产业、贸易、股权、担保等数据，围绕行内外企业构建关联网。一是建设产供应链图谱，挖掘企业的贸易特征，判定企业的市场地位、经营状态、行业生态等信息，辅助业务人员从宏观产业发展及微观企业多个视角全面评价企业。二是利用图分析能力支持联动营销，找到与目标企业关联的客户作为营销突破口，提供资金往来触达、间接股权触达等营销路径。三是通过建立产品匹配模型，为不同市场地位、不同经营状态的企业匹配特定的金融产品，支撑贸易金融、科技金融业务营销决策，提升精准营销效率。

后续，交通银行将进一步聚焦降成本、控风险、优体验、增效益四大目标，推动场景建设落地，扩大图谱应用规模，为业务高质量发展赋能。

## 民生银行知识图谱建设与应用

中国民生银行数据管理部 周礼 沙利民 张翔 夏玮璐

知识图谱作为前沿创新技术，可以融合银行内部和外部数据，从原始数据中深度挖掘各类实体之间的关联关系，进而识别企业风险、异常担保、疑似实控人等情况，为银行的风险防控、营销拓客等业务提供决策支持，更大程度地挖掘数据背后的价值。为此，民生银行建设包含“1+2+N”模式的“万象”知识图谱体系。以分布式图数据库与图计算框架为核心，构建了支持知识工程全流程处理的一站式知识图谱平台；实现了面向分析和面向交易两类图谱能力，分别用于关联挖掘与实时组网决策；融合行内外数据，构建了“N+”金融领域图谱，以超结构化数据驱动风险、营销、审计等数十个场景的业务创新。

### 打造一站式平台，支持分析与交易

根据对知识工程和复杂网络分析的能力抽象，基于 Hadoop/Spark 技术栈、整合图数据库、图算法以及知识构建工作流等相关组件，构建了“万象”知识图谱平台，提供了知识图谱建模、图谱分析、图谱挖掘、知识服务全流程的一站式应用能力。民生知识图谱平台主要包含四层内容，具体如下。

**1. 基础设施层。**在存储方面，以图数据库为核心，构建了多级存储机制。实体关系数据利用图数据库进行存储，并实现了生产多模态集群，分别服务于分析场景和交易场景；用于查询加速的索引数据以及关系详情等数据，使用 ES 或 HBase 等组件进行存储。在算法挖掘方面，通过图数据库以及 Spark GraphX 等计算框架的支持，提供对常规的图遍历、中心性、社区算法，以及图神经网络等深度学习算法的支撑，并且将常用算法封装为算子，便于研发人员使用。

**2. 图谱工程层。**为降低研发门槛，平台提供了界面化的工程能力。针对图谱研发的全流程提供了各类组件，能够支持各类大数据平台的数据整合和知识抽取，实现可见即所得的本体设计，并能够通过“拖拉拽”的挖掘流程构建复杂图谱，以及基于模型算子进行关系挖掘，形成的数据内容还可以一键发布为引擎服务。将这样的一站式能力供给行内各团队使用，形成“平台即服务”，从而推进图谱数据和应用场景的规模化落地。

**3. 图谱服务层。**为同时满足面向分析和面向交易的服务模式，平台在服务层构建了图谱分析模块和图谱服务引擎模块。其中图谱分析模块主要面向数据分析师和业务分析人员，提供交互式的知识图谱分析和探索能力；支持在各类图谱上进行自助查询、路径分析、群体分析、形态匹配等，并可以对分析结果进行快照导出或者文件下发进行进一步使用。而图谱服务引擎模块则将图谱数据或者图计算能力封装形成 API，对接业务系统形成实时在线服务。针对已建设好的实体、关系或者图数据，以服务对接或者嵌入式图谱等方式供业务系统使用。而针对交易决策场景，可以提供实时组网类或网络特征计算类服务，接受交易请求数据进行同步网络构建，根据策略计算各类特征并实时返回判定结果，嵌入交易流程形成智能决策支持。

**4. 业务产品层。**将图谱的各项数据、服务和能力封装后，可以形成独立的数据应用产品，如对标互联网企业图谱产品的行内金融图谱，覆盖了全国工商企业的档案和各类关系，并与行内客户数据实现了拉通，可以提供零售客户、法人客户和集团客户的全景视图。

## 构建多领域图谱，赋能数字化转型

民生银行持续夯实图谱内容与落地业务场景，以广泛支持各业务的数字化转型工作。在数据层面，通过不断引入工商、司法、税务、舆情等数据源和粗加工关系，与行内的交易、客户、风险、担保、申请、设备等数据进行拉通聚合，形成了企业图谱、信贷风险图谱、集团图谱、资金交易图谱等多项图谱资产；在赋能层面，基于平台和数据能力，当前在客户洞察、风险防控、营销获客、合规预警、经营管理等五大方向等 20+ 业务领域进行了场景实施。

**1. 构建企业图谱，赋能数字化风控与营销场景。**由于企业跨地域、集团化经营业态的发展，使得客户风险表现形式更加多元化，信息不对称进一步加剧，风险管理难度持续加大。为了应对复杂的经济形势，加强数字化风控能力，民生银行扩大观察视角，从客户整体关系网络角度评判风险。同时在营销场景中，传统的营销方式已无法满足快速变化的市场要求，民生银行顺应科技浪潮，占领发展先机，大力发展数字化营销。

民生银行首先实现行内客户信息、风险相关数据和行外工商数据的拉通，然后提取法人、自然人、集团三类实体，和担保、抵质押、股权、法定代表人、额度使用等几十种基础关系，并通过图算法深度挖掘担保圈、担保链、互保、集中性担保、间接持股等隐形关联关系，进而构建企业全息图谱，共计近百亿数据；并对复杂的关系图谱网络进行可视化建设，根据不同关系类型提供不同可视化方案，做到企业客户“一眼清”，为风险管理全流程中各类业务人员、管理人员以及各类风险管理系统提供关联分析支持。

同时企业图谱也应用于营销获客场景中，实现基于客户关系网络的营销线索挖掘，筛选出与



我行存量客户有关联关系且关联关系符合一定特征的潜力客户，下发经营机构组织开展精准营销及客户关系维护，经统计此类潜力客户的营销成功率是无关系客户的 6 倍以上。

**2. 构建交易图谱，全方位保障资金交易安全合规。**将资金交易抽象为网络，将一次交易作为网络中的边，交易双方为网络中的点，每一条交易边从资金转出的一方起始，指向资金转入的一方。但是交易网络的构建与计算技术难度较大，首先是网络中的点、边数量都非常多，且随时间推移只增不减，这与相对稳定的企业集团关系是不同的。其次属性信息量也很多：点属性涉及客户的身份信息、账户信息；边属性涉及交易时间、金额、附言、渠道等。再次，图谱中会存在明显的超点，即单位时间内交易次数明显高于其他账户的账户实体。超点往往会极大影响图计算的性能。

但交易图谱的应用范围非常广泛，包括异常资金流向、交易环路等都以它为基础。票据贴现资金流向监控是典型应用案例之一：央行的 126 号文等监管文件对于票据贴现业务给出了明确的监管要求，严禁票据贴现资金回流前手，严禁以贷款、贴现资金作保证金办理银行承兑汇票，虚增存款。以人工跟踪贴现资金流向效率较低且监控范围有限，亟需进行更高效、更全面的流向预警手段。

在高性能的图存储和计算能力基础上，结合图算法解决了上述技术难题，也为资金流向监控提供了新方案。我们以交易流水、票据信息、背书历史、客户信息作为原数据，抽取出票据、公司、账户六类实体，和出票、收款、背书、贴现、转账、所属六种关系，进而构建形成票据的交易网络图谱；运用图算法进行社团划分与流向传播，识别交易网络中环路、汇聚、群组等复杂交易形态；运用流水中交易对手、资金用途、交易附言进行预警规则研判；运用机器学习算法预测贴现资金违规概率；最终形成了覆盖同名划转行外、流入保证金、流入房市、流入理财、流入股市、回流前手、流入存单七类违规场景的预警信息，并提供违规案例的可解释性图谱，便捷辅助业务决策。

**3. 构建信贷申请图谱，实现全流程实时组网风控。**近年来欺诈风险呈现多产业链、多团伙等特点，团体欺诈风险识别难度愈加递增，传统的反欺诈体系已无法满足当前复杂欺诈模式的识别。同时信贷领域愈加强调掌握用户间的关联关系以及信贷申请行为的聚集性特征，而知识图谱在识别用户关系和聚集性特征方面就有着不可替代的优势。民生银行为更好地保护客户资产安全，增强自身欺诈识别能力，通过知识图谱建设了信贷全流程的实时组网风控体系。

通过知识图谱平台将信贷申请信息、设备信息、联系方式、工作单位、地址、受托支付、交易等行内关系数据和外部工商数据加工成信贷申请图谱，同时应用图实时计算技术，打造从关系视角进行风险分析的实时组网风控体系，支持在贷前、贷中、贷后全流程识别关联风险。并将实时组网风控结果进行可视化展示，提供交互式分析功能，配合相关的特征指标，帮助业务核查风险。目前知识图谱已为小微和消贷的 50+ 信贷产品申请提供实时组网与风险计算服务，响应时间达到毫秒级别。

## 图数据库建设发展思考与实践

平安银行金融科技部云数据运营中心 陈灿荣

随着 Oracle、MySQL 等关系型数据库的使用范围不断扩大，慢慢的暴露出一些使用场景上无法解决的问题。同时，互联网的发展也对传统数据库产生了很多不适应，过去几年也出现了大量的新型 NoSQL 数据库，图数据库就是其中之一。它的数据模型主要以边和节点来体现，相比传统数据库它的优点是可以快速解决复杂的关系问题，是能直观表达真实世界的一种数据库。基于以上背景，平安银行引入了 Neo4j 作为图数据库的选型，目前主要应用到以下几个场景。

**一是银行 CMDB 系统。**CMDB 包含了各种配置项全生命周期的信息以及配置项之间的关系 (包括物理关系、实时通信关系、非实时通信关系和依赖关系)，是天然适合图数据库的使用场景。假如只是存储资产，配置信息这些基础功能传统数据库就可以满足，但是 CMDB 的资产多，而且不同资产配置的属性，标签都不一样，放在传统数据库里表的种类、字段也各不相同，所以这类数据我们选择放在 MongoDB 里面，比起 MySQL 等传统数据库，MongoDB 更为灵活。基础配置数据有了，最大的问题就是如何展示各类配置之间的关联关系。因为每个配置之间的关系不是孤立的，想用起来就必须关联起来，假如放在传统数据库，可以通过主外键、表关联查询出结果，但是量多、关联层级也高的情况下性能表现不尽如人意。举个例子，假如我要查一台机器的主机名对应的 IP 是多少，在库里可能很容易查出来，但是我再想查它运行在哪台物理机上？然后我还想看这台机器接入了哪台交换机，还可以沿着这条线一直查关联关系，在传统数据库上层次越多，关联效率就越慢，但是在图数据库通过节点的属性过滤和边的关联查询，就可以很轻松且直观的把整个链路的数据展示出来，所以我们将 MongoDB 的数据导入到 Neo4j 中，利用图数据库的特性去做数据展示和关系分析。

**二是行内应用关联分析系统。**众所周知，各应用之间的调用链路是错综复杂的，想要获得和维护这份数据极具挑战，但是这项数据对于我们故障排查问题以及业务影响分析又是相当有用的。试想一下，我们日常维护一个应用可能只会通知第一下游关联方，但是还会有更多级联的关联方是我们忽略的对象，对他们可能也存在业务影响。或者在生产排障的时候，怎样才能迅速关联出两个甚至多个应用存在关联影响？所以行内也针对应用之间的访问关系进行采集，并且将数据存

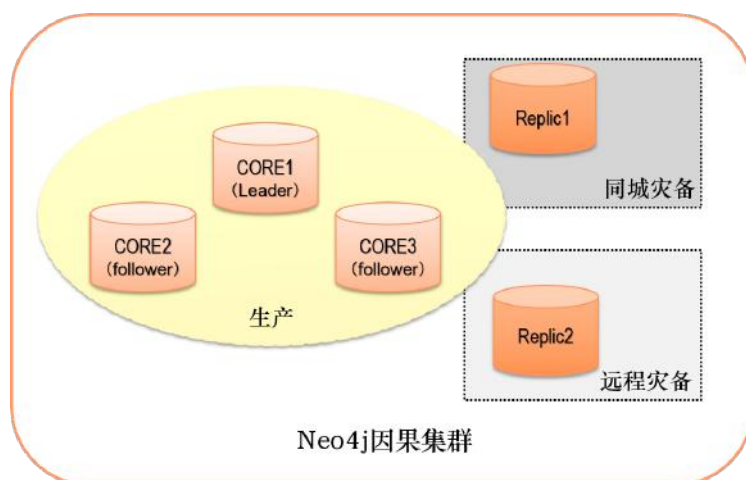
入图数据库，利用图库对关系的运算能力，可以轻松查询出各应用之间的关联关系。除此之外，对公反欺诈风控系统等也采用了图数据库。

银行的图数据库选型建设中，我们选型了 Neo4j。首先 Neo4j 作为现在业界最流行的数据库，并且有开源版本，文档的丰富度以及社区的活跃度都是行业顶尖的；其次提供的算法也很丰富，如 APOC、GDS 作为插件可提供很多便利的功能，而且高可用、运维便利性等综合性能较高。下图是我们 Neo4j 的一个架构，生产环境采用 3 个 CORE 节点组成保证高可用提供对外服务，同城和远程采用 Read\_Replica 方式配置只读副本，同城可作为只读节点对外提供读访问，还可作为跑 GDS 插件算法使用的节点。

但是图数据库在日常使用中也存在痛点。例如跟传统数据库之间缺少联动的平台工具，如 Oracle 的数据导入 MySQL 中，虽然是两种类型数据库，但是业内有很多成熟的方案进行数据同步，例如 ETL 工具，或者 GoldenGate 等，但是这类传统的抽取转换平台工具无法很好地兼容图数据库，很多情况下更依赖项目的开发方自行开发程序处理数据的导入初始化。所以选型 Neo4j 作为图数据库也是其中一个原因，Neo4j 提供很多便利的插件类似 APOC，可以直接通过 mysqljar 包连接到 MySQL 库将查询结果导入到 Neo4j 中，非常方便。还有 Neo4j 可通过各种插件与 Elasticsearch、Kafka 等相结合使用。

在处理错综复杂的数据的场景下，图数据库可以给人们以直观的数据展示，很好地弥补传统数据库薄弱之处，在深度数据处理、关联查询场景相比还有着优秀的性能表现。针对图数据库，我们可以对未来做一些展望，第一个是类似 SQL 标准的统一查询语言，现在的每款图数据库产品的查询语言不尽相同，

像 Cypher、Gremlin 等，统一的开发语言可以大大的降低学习成本，降低图数据库的学习入门的门槛。第二个希望有优秀的国产图数据库的出现。第三个是有流行可靠的开源工具平台可以便利地实现关系型数据库迁移到图数据库，这样将会更有力地将图数据库在金融业进行推广。





## 基于图数据库的金融知识图谱典型应用

兴业数字金融服务(上海)股份有限公司数据服务中心 王敏

### 建设背景

知识图谱作为人工智能技术的重要研究方向与组成部分,正在快速渗透到金融业务应用层面,并日益成为金融科技的“底层支撑”。随着近年来金融数据的爆发式增长,传统风控系统逐渐无法满足业务需求,而知识图谱技术利用在风险识别能力和大规模运算方面的突出优势,通过创建本体模型,根据实体间关系形成关联数据网的图谱形式,描画个人、企业等各类实体的基础信息、金融行为、关系网络等综合画像,串联大量多源异构数据,挖掘数据深层价值。

### 业务痛点

在传统的风控运营中,尽管传统的专家规则和机器学习模型在减少欺诈损失方面起到了很重要的作用,但随着科技的发展和黑产产业链的日趋成熟,欺诈手法也在不断升级,有组织的、大规模的、团体性的作案方式屡见不鲜,传统的风险防控手段也无法及时、准确地评估客户风险,无法有效应对新型的欺诈手段。引入知识图谱技术,使兴业数金初步实现了在观察视角上从“个人维度”到“全局关系”的转变,反欺诈手段由“由点到面”的提升,个体隐藏关联欺诈风险“由浅到深”的挖掘,为银行数字化转型建设提供重要基础。

### 图数据库实施路线

**1. 图数据库选型。**兴业数金结合自身技术体系及业务场景特点,基于对图数据库的基础数据服务能力、可视化管理功能以及性能等多方面进行考量,通过对比验证后发现,Nebula 图数据库在多跳遍历表现较好,可满足实时图指标计算、图挖掘等要求,最终选择了高性能、高可用、分布式可扩展的 NebulaGraph(后简称 Nebula) 图数据库作为应用技术底座,适配不同场景的存储及计算需求,并进行具体应用层的平台功能实现。

**2. 平台规划与建设。**兴业数金遵循图计算服务平台的基本能力要求,打造了一套全行级的知

识图谱平台应用体系，如下图所示，基于图数据库的底座支撑，搭建图计算及图谱应用平台，构建关联关系网络，以可视化技术，图算法为基石，提供全方位的图设计、图搜索、图分析、图规则、团伙识别、图探索等功能，满足行内各业务条线对于关联关系特征、规则及模型的快速开发及优化需求，提供一站式图构建、图计算和图应用等服务开放能力，以支撑兴业数金各业务场景应用，帮助业务快速发现风险点、可疑团伙和关联关系，实现“人机结合”形式的智能化分析、研判和决策，有效提升了风险防控及管控水平。

(1) “1+N”模式，一个平台支撑多个场景，降低图谱构建和运维成本。通过构建全行通用统一的知识图谱平台，实现各业务条线的场景图谱的集中管理和知识复用。平台通过低代码、配置化方式实现数据接入、实体关系抽取和图谱更新，并支持批量构建、实时更新多种场景，高效完成图谱构建与更新。同时，提供任务调度管理、图谱评价等运维功能，帮助运维人员快速识别图谱中异常数据及任务。

(2) 统一存储配置中心，兼容适配多种图数据库。平台采用组件化、插件化的可插拔架构设计，通过对图查询和图计算的抽象层设计，可快速兼容适配多种图库如 JanusGraph、Nebula 等，各场景图谱也可独立配置其特有的存储方案，以适应不同场景图谱对图数据库的个性化需求，实现各图谱的数据权限安全隔离。

(3) 构建分布式图挖掘引擎，支撑百亿级关系图分析挖掘。基于分析挖掘的深度应用需求，平台内置子图抽取、关系推理补全、群组挖掘、图指标计算等各类图挖掘函数及任务模板，同时包含几十种通用、自研高阶图算法及特定场景图挖掘模型，可实现在百亿级关系图规模下完成各



▲ 图 兴业数金全行级知识图谱平台应用体系

类复杂图挖掘任务，有效支撑行内风险传导分析、团伙挖掘等场景。

**3. 应用场景——信贷申请反欺诈。**基于信贷申请的进件数据、客户数据、人行征信数据以及逾期标签、历史客户贷后表现等数据，再结合外部的三方数据如多头借贷数据、外部黑名单、外部欺诈评分数据等，通过数据清洗加工、融合处理，然后再利用知识图谱平台进行知识抽取，构建信贷申请反欺诈关联图谱。以可视化关联分析工具形式，集成到审批系统和风控决策平台，供调查人员进行关系圈的洞察和匹配关联，发现中介包装、虚假进件等聚集和交叉关联模式和结构，为业务决策分析提供工具支持。同时，基于知识图谱关联指标，构建团伙识别模型、关联规则策略，共同为信贷事中检测和事后分析的业务赋能，进一步提升反欺诈调查效率和风险感知侦测能力。

基于所构建的信贷申请反欺诈图谱，以关系拓展和挖掘为核心，对实体及其关系进行直观的可视化展示，同时结合图计算引擎为用户提供多维度、多层次的分析模式，支持单一个体的画像、关联路径分析到群体的集中性识别、关系演变分析，帮助用户快速完成数据探索、关系挖掘、案件调查等任务，实现数据到智慧、智慧到决策的转变。

申请反欺诈通过引入知识图谱技术，大大提升了风控效果，不仅有效帮助业务部门识别出多种隐性风险，如中介包装、中介代办、虚假进件等聚集和交叉性关联风险模式，挖掘出数百个风险团伙，解决传统通过规则或模型手段无法快速识别的风险问题，还提高了业务人员的风险案件排查分析效能。

**4. 应用场景——企业关联关系图谱。**融合行内企金工商客户多源业务数据，如对公 CRM、ECIF、银行核心、对公信贷等数据源，并结合行外企业工商相关数据补充，进行数据清洗、知识抽取、知识加工、知识融合，构建全面的对公企业关系图谱，深入挖掘企业客户股权/投资关系、集团派系、企业上下游产业链、资金链、产品链上的金融服务需求，实现基于知识图谱技术的企业关联洞察、集团派系授信管控、关联风险识别、客户营销等业务应用。

通过融合行内外相关对公企业数据，建立企业 360 度全景视图，全方位展示企业工商信息、高管信息、上市信息、经营状况、关联关系、风险事件等等，实现企业客户的多维关联洞察。

基于所构建的企业基础关系，通过知识推理技术挖掘企业间的隐性关联关系，按照行内集团统一授信管理办法认定标准，利用图谱挖掘技术深入挖掘符合规则的全部企业，形成基于企业客观全貌的关联族谱。

基于行内客户抵质押物、连带保证等信息，全面梳理担保关系，挖掘普通担保、相互担保、循环担保等担保圈链，识别异常担保类型和担保链路，以全面掌握企业的担保风险状况，有效防范和化解担保圈贷款风险。

采集信贷系统、核心交易流水、贸易融资系统上下游企业信息，以群组的形式直观展示企业

产业链信息，分析企业在产业链中的地位及主营业务，发掘优质的企业客户，根据产业链客户需求营销供应链相关产品；分析上下游企业间的担保及应收应付账情况，根据行业特征及时推送风险及产品营销信息。

通过构建对公企业图谱，充分应用企业间的关联关系，有效挖掘潜在客户商机，捕获并及时防范隐匿的关联风险，全面提升存量客户价值，助力对公业务不断向前快速发展。

## 图数据库应用未来发展方向

离线场景大规模图数据的高效处理能力和实时场景高并发、低延时的图计算能力或是未来知识图谱建设的重点和难点。而图数据库对以上能力的建设起着至关重要的作用，图数据库大规模数据存储能力和查询能力以及实时图更新、实时图计算的能力自然是未来图数据库能力选型的关键所在。

另一方面，近年来以图神经网络为代表的图学习技术在人工智能领域掀起了研究热潮，图学习是一种人工智能相关的机器学习方法，通过学习节点和边之间的关系，从图数据中提取有用的特征和模式，可用于预测、分类、聚类和异常检测等任务。图学习技术有助于解决过往机器学习算法可解释性低等问题，如何解决图学习在内存、硬件方面的瓶颈及大规模图深度学习的难题并应用于金融场景产生业务价值，是知识图谱未来探索的另一重要方向。

## 图数据应用建设探索与实践

渤海银行信息科技部

知识图谱技术作为深化人工智能应用的重要手段，能够深度挖掘数据背后的价值，成为商业银行近年来发展创新的利器。随着金融监管要求的不断提高、金融机构内部管理和经营迅速发展、数据体量的指数级增长，渤海银行对于客户、账户、员工、资金关系等关联分析的要求也越来越高，其中又以风险、内控等领域尤为急迫，而传统的关系型数据库已无法满足深层级、高效率的关系链路挖掘需求。面对新挑战，我行紧跟时代潮流，积极引用知识图谱技术，不断探索金融知识图谱在各领域中的应用与发展。初期，基于 Neo4j 图数据库进行应用探索，以需求为导向从应用场景出发先后构建了信用风险、内控、血缘关系三个独立的知识图谱，真正实现了图谱应用在多领域的首轮实践，取得良好效果的同时证明了知识图谱技术重要意义和可行性。总结经验后开启了图谱建设新阶段，引入 StellarDB 分布式图数据库，统筹算力共享资源；打通行内外数据、整合多路外部数据源保证数据一致性；分领域构建了全量企业、信用风险、内部管理三个领域图谱；结合图计算和机器学习技术，赋能对公业务、内部管理、风险管控、内控合规等多领域业务，不断挖掘金融知识图谱技术带来的新价值。

### 案例一：内控合规场景

近年来，各个银行分支机构内部合规风险案件频发，个案涉及金额动辄上亿元甚至上百亿元。针对不断发生的风险事件，监管机构更是多次下发制度文件以及给予行政处罚，规范和防控各类违规事件。当前常规的合规监测方法主要有人工质检、专家规则与算法模型，但上述监测方法都具备一定程度的局限。人工质检核实违规行为的人工量大，频率较低，覆盖面少，并且无法进行标准化合规检查的推广；专家规则容易被人为探知，并且专家规则穷举耗尽，难以应对新模式的违规行为；而算法模型多度依赖先验经验，需要足量的黑样本数据，并且通过算法模型识别出的结果，很难给出逻辑解释。

面对日益复杂的合规风险场景，渤海银行通过引入知识图谱技术提升整体风控水平与决策引擎能力，使得监测结果兼具准确性与可解释性。首先，利用图计算和图挖掘模型，对冗长、复杂



的隐性关系进行深度的挖掘。其次，图谱计算出的关联特征指标可应用于专家决策分析，也可以为机器学习模型补充关联特征变量；最后，利用图谱的可视化，将模型命中情况、关系网络进行直观展现，满足对模型可解释性的要求。

实际落地中，渤海银行在票据中介团伙、对公资金异常流向以及员工疑似归集资金、理财飞单以及操纵客户账户等场景中，通过构建自然人（企业、个人客户或员工）、电话、设备、账户、集团 / 投资 / 担保、业务与交易等组成的关联图谱，利用子图挖掘、指标计算、图规则计算等手段，识别异常场景的有效率和关注率均达到 40%。以票据中介团伙为例，构建以主要交易对手、共用 IP/MAC、同一高管、同一集团、出票背书关系等多个维度上的团伙成员之间的关联图谱，从贴现笔数、贴现金额、贴现分行个数等多个维度解释团伙风险，能更有效的识别团伙风险，并且更清晰展示团伙成员之间的关联。

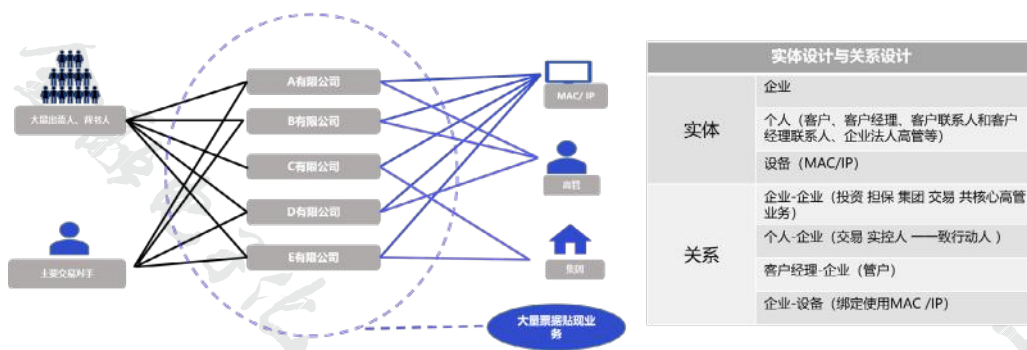
## 案例二：集团智能识别

因集团客户内部关联关系庞杂隐蔽，关联交易频繁且风险传导性强等特点，其识别管理成为银行信用管理领域的难题之一。目前渤海银行在集团管理上，识别依靠人工，且基于扁平化结构进行管理，无法识别成员间关系。基于此，渤海银行建设了集团智能识别模型进行风险识别。

一是针对行内 3000+ 集团，共计 11000+ 的行内客户，利用工商等外部数据结合深度有限遍历算法，补全了集团成员间的层级关系。与公司业务部共同对识别结果进行核实。针对 8000 万左右工商企业，利用数据加图谱算法，共识别约 200 万个集团，自动出具认定结果辅助集团认定管理。

二是完善集团客户管理的信息化手段，结合使用体验，将扁平化、清单化的集团管理改为图形化集团图谱展示，减少管理人员认定集团的工作量，提升集团客户管理水平。

三是通过对集团授信信息、资产信息等进行整合，从整体层面对集团进行管控，识别集团整



▲ 图 图谱构建示意图 - 票据中介团伙

体风险。

### 案例三：账户预警模型

因尚未构建账户类风险预警规则，对于账户类风险未够成自动预警的护城河，风险识别依靠人工完成，账户类风险揭示能力较弱，基于此渤海银行建设了账户预警模型。针对全量授信客户，结合交易模型数据构建交易图谱，使用图模式匹配等相关图算法，识别检测 12 种监控异常账户行为（如借款企业资金存量下降）、24 种异常资金交易（如贷款、票据资金回流至借款人、保证人等），2022 年全年共识别 100+ 异常交易。同时将识别出的预警信号推送预警系统和渤银智脑，及时触达一线客户经理，进行信号展示以及预警处置。

### 案例四：数据资产血缘链路图谱

随着数据量不断增长，传统基于关系型数据库搭建的数据血缘已经无法满足使用，迫切需要重构一种更清晰、高效的方式，因此渤海银行构建了基于图数据库的数据资产血缘链路图谱。

数据血缘链路图谱应用于数据管理域，提升了数据管理工作的效率与价值释放。数据血缘图谱清晰展现数据资产“从哪来”“到哪去”，实现数据追根溯源，加强了可视化展现。通过解析元数据存储过程，梳理数据资产之间加工链路关系，利用图数据库技术构建血缘图谱，逐步构建完整的数据资产地图；提供不同层次的展现粒度，系统、表、字段，满足业务上不同应用场景的数据查询和辅助分析需要。

数据资产血缘图谱已投放至行内的数资管家平台，用户可以当前数据资产为中心点，关联系统、上下级影响、血缘关系、资产详情一点便知。

### 案例五：产业链知识图谱

银行对公业务逐渐由粗放式经营向数字化、精细化金融服务转型，同时结合业务发展规划，需围绕产业转型升级和全产业链抓行业客户。

在此背景下渤海银行构建了产业链全景图谱，基于多层级产品关系梳理产业链上中下游，并挂接企业，利用知识图谱技术搭建产业链全景视图。为产业链路查询，产品供销关系查询，细分产业项下营销白名单生成，产业政策传导等提供公共服务能力。

在未来很长一段时间内，渤海银行将从实际业务痛点出发，积极探索业务落地的场景，逐步在反欺诈、营销等领域，用好知识图谱这把利器。



## 基于知识图谱的智能风控平台实施案例

浙商银行金融科技部

在金融领域中，信用评估、反欺诈和风险控制都是最为关键的环节，近些年金融数据爆发式增长，传统风控系统逐渐力有不逮。传统风控普遍存在信息不对称、成本高、时效性差、效率低等问题。一方面传统风控针对非结构化数据的汇总难度大、人力成本较高并且效率低下；另一方面由于金融领域数据具有海量、多结构、多维度、价值大、时效高等特点，数据获取难度大，隐藏在数据中的深层次关系难挖掘。因此，解决银行风控的难点和痛点迫在眉睫。为解决金融领域数据孤岛、格式多样、信息不对等、数据关联关系难以挖掘等问题造成的风险管控难的问题，浙商银行综合应用图数据库、人工智能和大数据等技术构建基于知识图谱的智能风控平台，并广泛引入外部数据，整合银行内部业务和流程，制定覆盖全客户、涵盖全流程的智能风控方案，赋能行内风险控制和管理业务。

### 实施方案

浙商银行自 2018 年开始开展知识图谱技术的应用研究，引入图数据库相关技术，并以集团和企业间的关联关系为切入点，自主研发基于知识图谱的智能风控平台。该平台基于 Neo4j 图数据库，整合并存储了包含工商、司法、监管、征信等外部权威数据以及行内业务数据。并综合应用知识图谱、人工智能、大数据等技术挖掘数据关系，构建集团和企业关联关系挖掘模型。构建涵盖自然人、集团、企业、公司客户以及集团关系、投资关系、股权关系、担保关系、任职关系等多种实体和关系类型的金融知识图谱，实现对企业客户的全流程风险防控管理，为银行风控中客户准入、关联关系、客户画像、风险预警等功能提供了数据和技术支撑。平台架构如图所示，

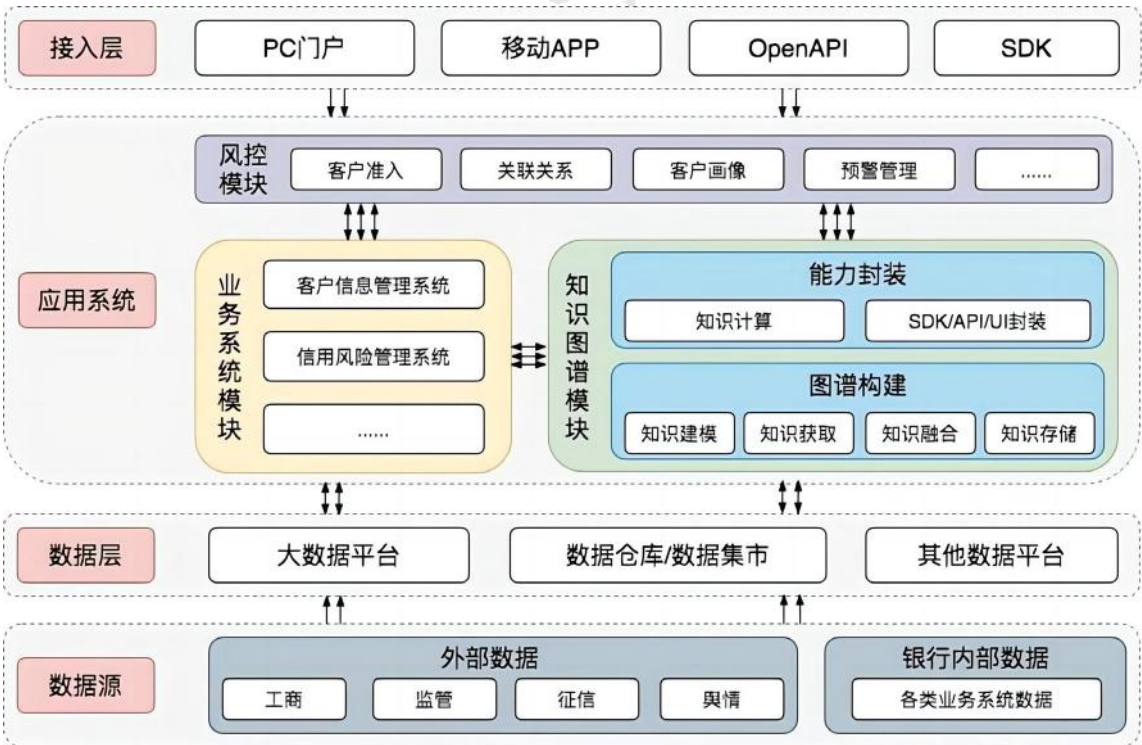
浙商银行基于知识图谱的智能风控平台按照数据层、应用系统层和接入层划分，数据层包括银行内部的大数据平台、数据仓库 / 数据集市和其他数据平台，运用大数据及人工智能技术完成各类数据的处理、加工；应用系统层包括集成的业务系统模块、知识图谱模块和风控模块，业务系统模块包括银行内部客户信息管理系统、信用风险管理系统等业务系统，知识图谱模块包括图谱构建模块和能力封装模块，该模块主要包括图数据库的设计和知识图谱的构建和存储；风控业

务模块提供了客户准入、关联关系、客户画像、预警管理等业务功能；接入层提供了 PC 端门户展示、移动端 APP 以及 OpenAPI、SDK 能力。

基于知识图谱的智能风控平台在图数据库设计和知识图谱构建过程中，首先对实体和关系进行建模，确定包含自然人和企业两类实体，集团和行内客户信息等其他类型分别以属性形式保存于两类实体中；综合利用工商、司法、监管和征信等外部数据和行内业务数据设计了任职、投资、持股、担保、集团等多种主要关系和行内认定集团、前十大股东等多种辅助查询的关系。并针对必要属性建立索引提升查询速度。在数据处理与入库存储过程中，针对多种数据类型分类处理，目前支持处理文本、表格、XML 和数据库数据等多种数据类型。

图数据库和知识图谱构建所涉及的关键技术包括但不限于：采用 OCR、NLP 等人工智能技术，针对银行内部包括财务报表、权证、票据等凭证的非结构化数据进行识别和提取；利用基于 HAWQ 的大数据分析引擎对工商、监管等外部数据源和银行业务数据进行统一的存储、加工和计算，并保存至图数据库中；平台采用基于关系数据库、Neo4j 图数据库和 ES 集群的混合存储结构，实现数据高效插入、更新和查询。

目前，浙商银行企业关联关系知识图谱规模达亿级，图数据库保存的实体和关系类型达 20 多种，支持基础查询、路径查询、关系推理等多种查询算法，并具备一定的推理和图计算能力。



▲ 图 基于知识图谱的智能风控平台架构图

## 应用成效

浙商银行基于知识图谱的智能风控平台 2018 年 11 月上线，截至目前已完成多次系统迭代升级。该平台自运行以来取得了良好的经济和社会效益。平台产生超过 900 万条预警信号，自动定级信号占比 95.98%，事前预警率达 82.35%，系统平均事前预警时间为 339 天，大幅提升了事前预警能力；该平台将大量人工操作转化为自动化系统流程，帮助银行运营管理降本增效，提升服务质量；同时，平台可协助业务人员深耕、细耕风险和收益综合性较好的企业客户，实现盈利增长。

## 未来发展方向

浙商银行未来将持续致力于图数据库、知识图谱、人工智能等创新技术赋能银行传统业务的智能化、数字化建设，后续将继续开展图数据库和知识图谱的研究和应用，一方面将不断提升图数据库的应用能力和知识图谱的构建能力，沉淀包括知识建模、知识获取、知识融合、知识计算等关键技术积累，推进知识图谱行业标准建设；另一方面将持续推进图数据库和知识图谱的场景化应用，包括利用图数据库对交易信息存储、分析和挖掘，构建交易知识图谱，结合图计算等技术提升行内反欺诈和反洗钱能力。也将进一步赋能信贷风控、营销获客、产品创新、运营管理等金融场景的智能化建设。

## 图计算在中原银行反欺诈领域的探索应用

中原银行数据信息部

### 项目简介

面对黑产欺诈和电信诈骗等活动日益猖獗且呈现高发多发态势，监管部门要求各金融机构采取有效风险防控措施，充分利用数据挖掘分析工具对账户异常开立和可疑交易情况进行实时监测，提高违规、欺诈交易识别能力。

具体落实中面临如下主要问题：一是金融欺诈呈现专业化、产业化、隐蔽化、场景化四大特征，应用传统的专家规则无法实现银行业务风险的精准防控。二是机器学习等算法通过概率学指标预测风险等级，其相关性往往不构成因果关系，造成业务可解释性较差，导致对欺诈行为举证困难。三是传统架构在处理多数据源、大数据量的关系分析场景时，响应速度无法保障，容易错失风险防控的黄金时间。

针对上述问题，中原银行利用图计算复杂网络分析能力强、大规模数据处理性能高、可视化展示效果好的特点，结合客户设备、账户、交易行为等数据，建立反欺诈模型，打造涵盖主体、行为、关系的多维智能反欺诈体系，实现对全行 14 个业务渠道、108 类业务场景的风险管控，全面提升业务风险的防范能力，保障客户财据安全，取得良好成效。

### 建设内容

中原银行基于图计算的智能反欺诈体系，通过图计算系统以图数据服务和图模型服务两种形式覆盖交易反欺诈、申请反欺诈、营销反欺诈三大业务方向，从事前、事中、事后风控管控闭环出发，实现涵盖主体、行为、关系的多维欺诈风险防控。

**1. 图数据服务建设。**图计算系统融合数据量达到 25.6 亿条，对外提供的数据服务包含 15 种实体，40 种关系，其中企业实体 2.1 亿余条，个人实体 5.3 亿余条，其他实体 1.5 亿余条，零售关系 4.25 亿余条，通用关系 7.39 亿余条，对公关系 5.1 亿余条。目前，图数据服务已形成基础标签、业务规则标签、智能标签三大类标签（见图 1），全面覆盖对公、零售、风险等业务条线，成为各部门数智化转型的基石。



▲ 图1 数据标签架构图

2. 图模型服务建设。面向不同场景特点提供端到端模型服务，已构建个人账户涉诈挖掘模型、对公涉赌涉诈模型、欺诈公司团伙图挖掘模型、紧急止付图挖掘模型、羊毛党挖掘模型等。

## 项目成果

1. 图数据服务方面。通过对接银行内业务系统，提供数据展示、决策辅助等功能，实现传统线上业务的智能化升级。

在对公场景中提供对公客户 360 视图，增强数据聚合服务能力。与对公 CRM 实现深度数据和功能融合，重点实现企业基本信息、关联关系、客户贡献度、已开展业务、行内动态、行外事件、拜访记录、舆情消息推送等企业全维度信息的展现。

在零售场景中接入多种数据源的零售数据，形成标准化的实体和关系数据，构建零售图计算模型。嵌入零售 CRM 系统，建立完备的权限体系，支持不同零售部门的功能和数据权限隔离。实现个人担保圈分析、非关联人贷款分析等场景在零售业务中的应用。

2. 图模型服务建设。截至目前，中原银行应用图计算系统协助地方公安机关破获欺诈案件 3 起，共发现可疑团伙 466 个，累计确认风险账户 18702 个，管控风险金额累计 3097 万元。

在交易反欺诈场景中，基于公安系统下发紧急支付账户名单构建紧急止付实时图挖掘模型，推送可疑线索 3501 个，管控账户 1138 个，涉及余额 729 万。通过欺诈公司团伙图挖掘模型发现 353 个可疑企业团伙，涉及 454 个可疑企业账户。



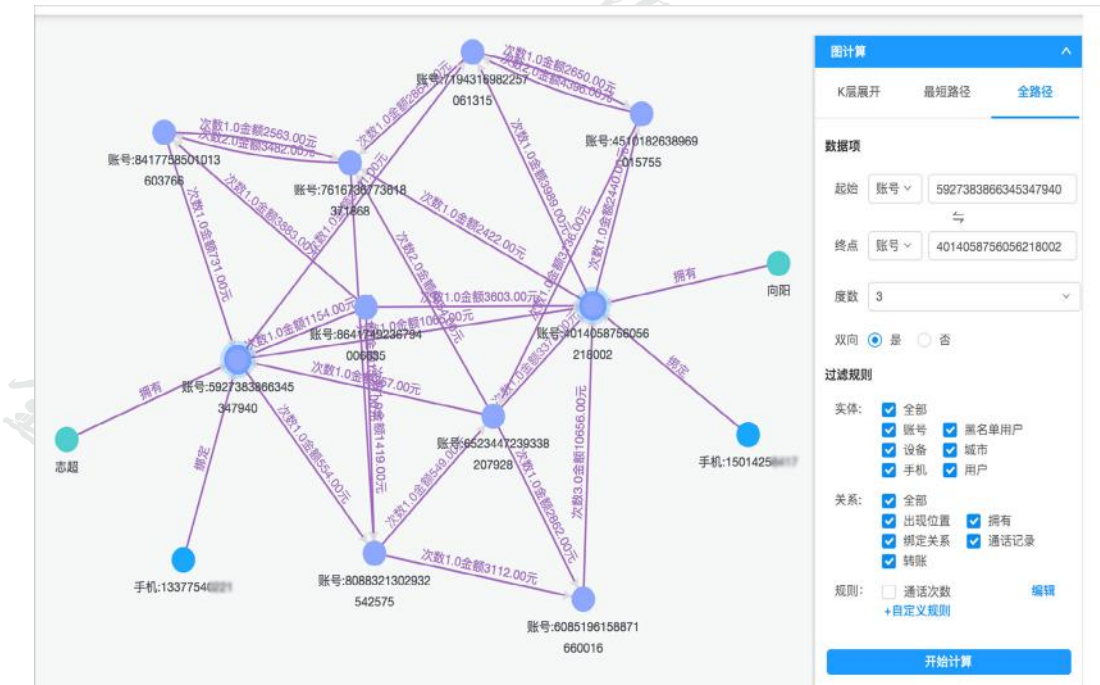
在申请防欺诈场景中，构建信贷业务申请欺诈团伙挖掘模型，共挖掘 19 万余个团体，精准筛选出近 200 个高危风险团体。

在营销反欺诈场景中，通过对参与营销活动的 62.4 万人进行分析，构建羊毛党挖掘模型，认定高危名单共计 70660 个，其中高危设备 83 个，高危 IP 58 个，高危手机号 70460 个，高危客户号 59 个，加入反欺诈黑名单库进行管控。

创新亮点

**1. 多维度图算法体系赋能精准防控。**中原银行提出“全渠道、全周期、智能化、实时化”目标，基于数仓和图数据库平台建立数据底座，融合基础算法与金融行业特色算法形成算法仓库，应用银行业务规则沉淀端对端模型，实现各条线全周期风险管理业务赋能。

**2. 多模态融合框架实现全周期覆盖。**中原银行采用 GraphX 图计算框架、NLP 关键字抽取等多种处理方式，对数据进行顶点和边的过滤、聚合、连接、映射，为后续的图模型挖掘和关系查询提供快速且充足的数据基础。借助 StarRocks + Flink + Kafka 完成实时数据的抽取与加工，实现分钟级、小时级、天级数据任务的多模态加工方案，为业务人员提供多种时效性的图数据可视化服务和图模型洞察服务。融合分布式缓存、微服务、容器云等技术，搭载高性能图数据库，具备多跳查询、关系强度计算、社区发现等场景的秒级计算能力。



▲ 图 2 转账关系图谱

**3. 可视化展示解决数据可解释性要求。**图计算系统从实体、属性、关系的角度将业务逻辑抽象为图计算模型，通过标签传播、图直径等维度形象描述欺诈团伙的形态和构成，并通过可视化方式链状、环状、网状关系，生成企业和个人的投资关系、担保关系、资金往来关系等关系图谱，业务人员能够从多维度、多视角了解客户历史变更及现状情况，有效提升精准营销和风险预警能力。（见图2）。

## 未来展望

未来，图计算技术在金融反欺诈领域的应用方向将更加多样化，将为金融反欺诈领域带来更准确、高效和可靠的解决方案。主要体现在：一是更高效的算法和模型。针对大规模的金融反欺诈场景，使用图卷积神经网络（GCN）或图注意力网络（GAT）等模型，对欺诈行为进行建模和预测，提高计算速度和准确度。二是更丰富的数据类型和数据源。除传统结构化数据外，需利用社交媒体数据等非结构化数据，更全面地了解客户行为和偏好，提高反欺诈的准确性。三是更好地保护隐私和数据安全。在应用图计算技术进行反欺诈分析时，结合隐私计算技术，在充分保护数据和隐私安全的前提下，实现数据价值的转化和释放。



## 阳光财险非车新一代产品工厂

阳光财产保险股份有限公司信息技术部

为支撑阳光财险未来 5 ~ 10 年的非车业务发展需求，提升保险服务水平，解决当前产品创新与管理的诸多痛点，阳光财险信息技术部、财产险部、意外险部，联合阳光保险集团科技中心，自主设计研发了阳光财险非车新一代产品工厂，并成功应用于阳光非车新一代承保和理赔系统建设中。

阳光财险非车新一代产品工厂基于数字化、低代码、云原生的设计理念，充分分析了非车业务特点，并借鉴了大量同业项目建设经验，探索性地将复杂的非车产品模型和组件关系等保险产品主数据基于图数据库维护，实现了产品配置化上线和分层管理能力，短时间内达到了服务提升、技术突破、业务赋能的目标，无论是在投产效果还是技术先进性上都达到了业内领先，有效提升了产品创新能力和保险服务水平。

### 背景和意义

保险作为社会保障体系的重要组成部分，是社会发展的“稳定器”和“助推器”，随着保险意识逐渐为大众所理解和接受，保险产品的快速创新能力和多元化配置能力显得愈发重要。非车险作为各保司重点发力的跑道，面临巨大的发展机遇的同时，也面临着竞争激烈、专业能力要求高的挑战，要在非车险市场获得长远发展，必须对传统非车业务转型升级。

阳光的非车险产品管理多年来存在产品市场响应速度慢、新产品配置同步各业务环节效率低、产品投放和上下架管理支撑弱等问题。为解决上述痛点，阳光非车新一代产品工厂在建设过程中遵循业务驱动第一的建设原则，借助科技力量和流程创新，实现产品在承保等全流程、全渠道的快速精准投放，意健险产品零代码配置上线，复杂产品上线效率提升 50% 以上。

### 项目内容

阳光财险非车新一代产品工厂系统基于阳光保险自研 SAAB 开发平台搭建，采用前后端分离的微服务架构，基于统一服务治理平台纳管，借鉴低代码、云原生的开发设计理念，将保险产品

主数据基于图数据库维护，实现视图设计、模型管理、产品配置、编排管理等主要能力，80% 以上的非车险产品可实现完全配置化上线，目前已提交发明专利申报。

产品工厂主要包括以下设计目标和特点：

**1. 满足保险产品相关视图配置化上线要求。**前端基于模型配置驱动提供“拖拉拽”形式的页面构建能力，所见即所得，可实现复杂前端视图组件 100% 配置化上线。同时自研契合保险场景的前端视图模型驱动，支持与后端数据模型绑定。

**2. 满足保险产品相关保单模型配置化上线要求。**后端基于 Neo4j 维护领域模型数据，支持保险条款、责任、标的、限额、特约等数据和规则的配置和计算公式配置，以元件、模型、组件、产品的理论模型层层驱动产品。借助产品工厂的配置能力，非车承保系统代码量由 900 万行缩减至 136 万行。

## 项目创新点与实施效果

**1. 创新点。**技术创新方面，阳光财险非车新一代产品工厂基于低代码设计理念，探索性地将复杂的非车产品模型和组件关系基于图形数据库维护，通过产品的原子化、组件化管理及多层次组装，可在短时间内配置出一款产品，实现产品对不同渠道、活动、服务和客群要求的快速响应，同时对多业务模式、差异化的承保和批改流程提供更好的支持。理念创新方面，实现了同质化产品向客户需求定制化解决方案的转变，具有敏捷创新、全领域同步、精准投放等产品管理能力，产品研发上线效率快速提升。流程创新方面，实现了基于产品工厂的线上化产品研发、多版本上线、下架全生命周期灵活管控，运营作业效率大幅提升。

**2. 项目实施效果。**经济效益方面，非车新一代产品工厂大幅节约了保险产品创新和运营成本。非车产品上线效率提高了 50%，每年可节约 4000 小时的操作时间，同时产品工厂对外提供标准接口，实现对于合作伙伴系统的快速对接，接入速度平均加快了近 2 周，协同沟通效率提升了 3 倍。系统自 2021 年 4 月试点上线以来，已推广至阳光财险全国 36 家二级机构，目前已覆盖 400 余款非车险产品，获得了公司内外的广泛好评。

非经济效益方面，阳光财险非车新一代产品工厂借鉴业内先进的科技理论基础，针对原有作业模式进行大胆革新和数字化转型，贯彻以客户为中心、场景驱动业务的设计理念，探索出一套适用于保险行业的低代码、云原生敏捷研发实践，提升了智能化、敏捷化、数字化、线上化、云化的五大技术端能力，推动了企业信息技术与业务共同变革，提升了保险服务水平。

获奖情况方面，本项目于 2023 年荣获中国人民银行评选的“2021 年度金融科技发展奖”。

## 参考文献

- 【1】 《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》
- 【2】 中国人民银行《金融科技发展规划（2022-2025 年）》
- 【3】 工业和信息化部《“十四五”软件和信息技术服务业发展规划》
- 【4】 中共中央、国务院《数字中国建设整体布局规划》
- 【5】 Gartner《2022 年图数据库管理系统市场指南》
- 【6】 赛迪顾问《2022 年中国图数据库市场研究报告》