



中国计算机学会  
CHINA COMPUTER FEDERATION



吉林大学  
JILIN UNIVERSITY

NDBC  
2025·吉林·长春

# (NDBC2025) 第42届CCF中国数据库学术会议

THE 42ND CCF NATIONAL DATABASE CONFERENCE (NDBC 2025)

会议手册

「2025年8月1日 ~ 8月3日」  
吉林省·长春市



# 目录

|                            |    |
|----------------------------|----|
| 1. 组织机构 .....              | 1  |
| 2. 征文通知 .....              | 2  |
| 2.1 征文范围（包括但不限于以下领域） ..... | 2  |
| 2.2 征文要求 .....             | 3  |
| 2.3 系统演示 .....             | 3  |
| 2.4 论文出版 .....             | 3  |
| 2.5 重要日期 .....             | 4  |
| 2.6 联系方式 .....             | 4  |
| 3. 系统演示征文通知 .....          | 5  |
| 3.1 征集范围 .....             | 5  |
| 3.2 投稿要求 .....             | 5  |
| 3.3 系统演示 .....             | 5  |
| 3.4 重要日期 .....             | 6  |
| 3.6 联系方式 .....             | 6  |
| 4. 注册通知 .....              | 7  |
| 4.1 注册费标准 .....            | 7  |
| 4.2 缴费方式 .....             | 7  |
| 4.3 CCF 关于会议注册费的退费规定 ..... | 7  |
| 5. 会议日程简表 .....            | 9  |
| 6. 研究生辅导报告 .....           | 12 |
| 6.1 论坛日程 .....             | 12 |
| 6.2 论坛简介 .....             | 12 |
| 6.3 论坛主席 .....             | 13 |
| 6.4 论坛报告 .....             | 14 |
| 问题驱动的科学研究 .....            | 14 |
| 研究生做科研的几点建议 .....          | 15 |
| 论文撰写与科学初探 .....            | 16 |
| 研究生科研入门——第一篇顶会论文怎么写？ ..... | 17 |
| 研究生科研：如何找论文与读论文 .....      | 18 |
| 7. 企业之夜 .....              | 19 |
| 7.1 论坛日程 .....             | 19 |

|  |    |
|--|----|
| 8. 大会报告 .....  | 20 |
| 8.1 日程安排 .....   | 20 |
| 8.2 大会论坛报告 .....   | 21 |
| 数据智能体：自主数据处理的新范式 .....   | 21 |
| 大模型：当前数据塑造智能的一条成功路径 .....  | 22 |
| 从应用视角看数据库发展与未来 .....   | 23 |
| Next-Generation Blockchains: Trends, Challenges, and Solutions ..... | 24 |
| Database Technologies for Healthcare .....                           | 25 |
| 9. 第五生产要素背景下的数据技术论坛 .....  | 27 |
| 9.1 论坛日程 .....   | 27 |
| 9.2 论坛简介 .....   | 27 |
| 9.3 论坛主席 .....   | 28 |
| 9.4 论坛报告 .....   | 29 |
| 对数据及其价值的再认识与再实现 .....  | 29 |
| 协同智能与通证经济：概述与前沿 .....  | 30 |
| 复杂系统数据治理的范式演进与探索 .....   | 31 |
| 一汽数字化转型中的数据技术 .....  | 32 |
| AI 智能体的法律治理路径：以最小必要原则为中心 .....                                       | 33 |
| 10. 大规模图数据管理与分析论坛 .....  | 34 |
| 10.1 论坛日程 .....  | 34 |
| 10.2 论坛简介 .....  | 34 |
| 10.3 论坛主席 .....  | 35 |
| 10.4 论坛报告 .....  | 36 |
| 高效大图数据近似计算 .....   | 36 |
| 大规模图数据上的正则路径查询 .....   | 37 |
| Nezha: 面向异构硬件的超大规模图处理系统 .....  | 38 |
| 图数据的高效压缩与查询 .....  | 39 |
| GOpt: 模块化统一图查询编译框架 .....   | 40 |
| 11. 时序数据智能管理与分析论坛 .....  | 41 |
| 11.1 论坛日程 .....  | 41 |
| 11.2 论坛简介 .....  | 41 |
| 11.3 论坛主席 .....  | 42 |
| 11.4 论坛报告 .....  | 43 |
| 时间序列大数据计算：探索与实践 .....  | 43 |

|   |    |
|---|----|
| 面向数据库自治服务的时间序列数据管理 .....                | 44 |
| 面向泛化与决策的下一代时空数据智能建模：跨尺度、跨域与跨智能体视角 ..... | 45 |
| 超时空图神经网络.....                           | 46 |
| 时空数据紧致无监督学习与预测技术.....                   | 47 |
| 基于时序智能的云平台智能巡检.....                     | 48 |
| 12. 数据库系统新技术论坛 .....                    | 49 |
| 12.1 论坛日程.....                          | 49 |
| 12.2 论坛简介 .....                         | 49 |
| 12.3 论坛主席 .....                         | 50 |
| 12.4 论坛报告 .....                         | 51 |
| 神经符号数据库技术及应用 .....                      | 51 |
| 大模型时代的数据管理技术：进展与展望 .....                | 52 |
| 三层解耦云原生数据库关键挑战与技术创新 .....               | 53 |
| 数据库研究中的一些冷门却仍然重要的问题 .....               | 54 |
| 大模型赋能的数据库自优化技术 .....                    | 55 |
| 12.5 Panel .....                        | 56 |
| 数据库系统前沿技术与智能化演进 .....                   | 56 |
| 13. CCF-华为胡杨林基金数据库专项论坛 .....            | 57 |
| 13.1 论坛日程 .....                         | 57 |
| 13.2 论坛简介 .....                         | 58 |
| 13.3 论坛主席 .....                         | 59 |
| 13.4 论坛报告 .....                         | 60 |
| 基于对象存储的存算分离数据库数据结构 .....                | 60 |
| 通用解耦 OLTP 优化器的高效描述接口 .....              | 61 |
| 查询资源预估和智能任务调度 .....                     | 62 |
| 面向场景化机器学习任务的高效大数据存储格式设计 .....           | 63 |
| 面向 NPU 异构计算资源的软硬协同调度算法 .....            | 64 |
| 结构与非结构化数据关联知识发掘 .....                   | 65 |
| 结构与非结构化数据关联知识发掘 .....                   | 66 |
| 基于大语言模型和轻量模型结合的多模态数据分析 .....            | 67 |
| 基于大语言模型和轻量模型结合的多模态数据分析 .....            | 68 |
| 统一多模态数据融合检索技术 .....                     | 69 |
| 统一多模态数据融合检索技术 .....                     | 70 |
| 统一多模态数据融合检索技术 .....                     | 71 |

|  |    |
|--|----|
| <b>14. CCF-阿里云瑶池科研基金论坛 .....</b>         | 72 |
| <b>14.1 论坛日程.....</b>                    | 72 |
| <b>14.2 论坛简介.....</b>                    | 72 |
| <b>14.3 论坛主席.....</b>                    | 73 |
| <b>14.4 论坛主持人.....</b>                   | 73 |
| <b>14.5 论坛报告 .....</b>                   | 74 |
| 开场致辞、基金工作整体汇报.....                       | 74 |
| 面向 CXL 内存池的近数据计算架构与优化.....               | 74 |
| 基于 CXL 技术的云原生数据库架构设计与优化.....             | 75 |
| 软硬结合压缩数据直接计算数据库系统研发 .....                | 76 |
| 基于 GPU 和 NVMe-SSD 的软硬协同向量数据管理关键技术研究..... | 77 |
| SysInsight：基于代码大模型的数据库系统参数调优 .....       | 78 |
| 面向车联网的海量多模数据压缩技术.....                    | 79 |
| 数据库测试技术研究.....                           | 80 |
| OLAP 的多租任务隔离与运行资源优化研究 .....              | 81 |
| 论坛总结及致谢.....                             | 81 |
| <b>15. CCF-蚂蚁科研基金数据库专项论坛 .....</b>       | 82 |
| <b>15.1 论坛日程.....</b>                    | 82 |
| <b>15.2 论坛简介 .....</b>                   | 82 |
| <b>15.3 论坛主席 .....</b>                   | 83 |
| <b>15.4 论坛报告 .....</b>                   | 84 |
| 基于 CPU-GPU 异构计算架构技术研究.....               | 84 |
| 云原生多维效能感知的智能缓存技术.....                    | 85 |
| 高效大容量向量数据库引擎的研究 .....                    | 86 |
| 数据库技术与视频多模态大模型深度融合技术研究 .....             | 87 |
| Text2SQL 复杂语义理解与推理研究.....                | 88 |
| 基于数据库的检索增强生成研究 .....                     | 89 |
| <b>16. 软件学报专刊论坛 .....</b>                | 90 |
| <b>16.1 论坛日程.....</b>                    | 90 |
| <b>16.2 论坛简介 .....</b>                   | 90 |
| <b>16.3 论坛主席 .....</b>                   | 91 |
| <b>17. 教育论坛 .....</b>                    | 92 |
| <b>17.1 论坛日程.....</b>                    | 92 |
| <b>17.2 论坛简介 .....</b>                   | 92 |

|   |     |
|---|-----|
| 17.3 论坛主席 .....   | 93  |
| 17.4 论坛报告 .....   | 94  |
| 中国人民大学数据库系统课程的传承与创新 .....   | 94  |
| 语言教育科技：智慧教育的新问题 .....   | 95  |
| 面向新一代信息技术的山东省计算机科学与技术专业建设：融合与创新 .....   | 96  |
| 计算机学科数据库教育体系建设实践与探索 .....   | 97  |
| OceanBase 产学研人才培养 .....   | 98  |
| 17.5 Panel .....  | 99  |
| 大模型时代下数据库系统课程设计的机遇与挑战 .....   | 99  |
| 18. 贝叶斯数据分析论坛 .....   | 100 |
| 18.1 论坛日程 .....   | 100 |
| 18.2 论坛简介 .....   | 100 |
| 18.3 论坛主席 .....   | 101 |
| 18.4 论坛报告 .....   | 102 |
| 贝叶斯决策理论引导的机器学习 .....  | 102 |
| 结构信息的智能感知与博弈新理论及方法 .....  | 103 |
| 复杂数据偏差下因果推断及赋能大模型思考 .....   | 104 |
| 信息论视角下的贝叶斯深度学习与图分析技术 .....  | 105 |
| 18.5 Panel .....  | 105 |
| AI 时代的贝叶斯数据分析技术现状、挑战与未来 .....   | 105 |
| 19. 新一代数据库与数据赋能技术国际论坛 .....   | 106 |
| 19.1 论坛日程 .....   | 106 |
| 19.2 论坛简介 .....   | 107 |
| 19.3 论坛主席 .....   | 107 |
| 19.4 论坛报告 .....   | 108 |
| The Destiny of Database is Data Power Platform .....  | 108 |
| AI and Smart Cities: Advanced Technologies for Safe and Comfortable Mobility ....                 | 109 |
| Big Data Systems for ML .....   | 110 |
| The Acceleration of Graph Learning based on Ascend NPU .....                                      | 111 |
| A Content Recommendation Method Based on Future User Profiles to Support Interest Expansion ..... | 112 |
| OceanBase: The Fastest Distributed SQL Database for OLTP, OLAP and AI .....                       | 113 |
| Industry Intelligence Driven by the Dual Wheels of Data and Knowledge .....                       | 114 |
| MoE-Enhanced Personalized Federated Learning .....  | 115 |
| Privacy in Fine-tuning LLM: Attacks, Defenses, and Future Directions .....                        | 116 |

|   |     |
|---|-----|
| <b>19.5 Panel .....</b>   | 117 |
| <b>The Future of Multimodal Data Management: At the Crossroads of Technological Breakthroughs and Ecosystem Collaboration .....</b> | 117 |
| <b>20.数据库开源技术论坛.....</b>  | 118 |
| <b>20.1 论坛日程.....</b>   | 118 |
| <b>20.2 论坛简介 .....</b>  | 118 |
| <b>20.3 论坛主席 .....</b>  | 119 |
| <b>20.4 论坛报告 .....</b>  | 120 |
| <b>数据库的开源协同创新发展：以 Apache IoTDB 为例.....</b>  | 120 |
| <b>开源开放，生态共赢——OceanBase 开源社区技术演进与 AI 融合实践 .....</b>   | 121 |
| <b>A Unified Narrative for Query Processing in Graph Databases .....</b>  | 122 |
| <b>AI 时代的开源大数据发展现状、挑战和未来.....</b>   | 123 |
| <b>破土·生长·共生：PostgreSQL 在中国的演进之路、本土实践与未来展望 .....</b>   | 123 |
| <b>20.5 Panel .....</b>   | 124 |
| <b>数据库开源技术在中国的现状、挑战与未来 .....</b>  | 124 |
| <b>21. 移动云大云海山数据库合作生态论坛.....</b>  | 125 |
| <b>21.1 论坛日程.....</b>   | 125 |
| <b>21.2 论坛简介 .....</b>  | 125 |
| <b>21.3 论坛主席 .....</b>  | 126 |
| <b>21.4 论坛报告 .....</b>  | 127 |
| <b>时空数据库自然语言查询转换方法研究.....</b>   | 127 |
| <b>面向云原生数据库的 Serverless 数据压缩传输技术研究.....</b>   | 128 |
| <b>大云海山数据库产品与架构解析 .....</b>   | 129 |
| <b>面向新兴计算架构的高效图数据处理系统研究 .....</b>   | 130 |
| <b>如何在 K8s 上构建大规模的 DBaaS (Database as a Service) 平台 .....</b>   | 131 |
| <b>关键行业数据库云化的思考和实践 .....</b>  | 131 |
| <b>22. 萨师煊优秀学生论文汇报 .....</b>  | 132 |
| <b>22.1 报告日程 .....</b>  | 132 |
| <b>23. 分组报告 .....</b>   | 133 |
| <b>23.1 分组报告 1：数据智能与分析 .....</b>  | 133 |
| <b>23.2 分组报告 2：数据系统与应用 .....</b>  | 135 |
| <b>24. 系统演示 .....</b>   | 137 |
| <b>24.1 论坛日程 .....</b>  | 137 |
| <b>24.2 论坛主席 .....</b>  | 139 |
| <b>25. 赞助商风采 .....</b>  | 140 |

|                                |     |
|--------------------------------|-----|
| 25.1 赞助单位 .....                | 140 |
| 25.2 长期钻石合作单位: 华为 .....        | 141 |
| 25.3 长期钻石合作单位: 达梦数据 .....      | 145 |
| 25.4 长期铂金合作单位: OceanBase ..... | 149 |
| 25.5 长期铂金合作单位: 腾讯云数据库 .....    | 151 |
| 25.6 长期铂金合作单位: 金篆信科 .....      | 153 |
| 25.7 长期铂金合作单位: 阿里云 .....       | 155 |
| 25.8 长期铂金合作单位: TDengine .....  | 157 |
| 25.9 长期铂金合作单位: 金仓数据库 .....     | 159 |
| 25.10 长期铂金合作单位: 南大通用 .....     | 161 |
| 25.11 长期铂金合作单位: 移动云 .....      | 163 |
| 25.12 长期金牌合作单位: 海量数据 .....     | 165 |
| 25.13 长期金牌合作单位: 百度 .....       | 166 |
| 25.14 长期金牌合作单位: 字节跳动 .....     | 167 |
| 25.15 长期金牌合作单位: 崖山数据 .....     | 168 |
| 26. 承办单位介绍 .....               | 169 |
| 27. 会务信息 .....                 | 171 |
| 27.1 会议会场 .....                | 171 |
| 27.2 会议交通及住宿 .....             | 171 |
| 27.2.1 乘车指南 .....              | 171 |
| 27.2.2 住宿环境 .....              | 173 |

## ■ 1. 组织机构

**主办单位：**中国计算机学会

**承办单位：**中国计算机学会数据库专业委员会、吉林大学

**大会主席（General Co-Chairs）**

马琰铭（吉林大学）

陈刚（浙江大学）

**程序委员会主席（PC Co-Chairs）**

童咏昕（北京航空航天大学）

李博涵（南京航空航天大学）

**系统演示主席（Demo Co-Chairs）**

赵翔（国防科技大学）

徐辰（华东师范大学）

**执行主席（Executive Chairs）**

杨博（吉林大学）

**组织委员会主席（Local Co-Chair）**

申海（吉林大学）

张小利（吉林大学）

**论坛主席（Session Co-Chairs）**

王鑫（天津大学）

柴云鹏（中国人民大学）

屠要峰（中兴通讯研究院）

**宣传主席（Publicity Chairs）**

王宏志（哈尔滨工业大学）

**出版主席（Publication Co-Chairs）**

崔立真（山东大学）

张岩峰（东北大学）

**志愿者主席（Volunteer Co-Chairs）**

李晖（贵州大学）

李婷婷（吉林大学）

## 2. 征文通知

CCF 中国数据库学术会议始于 1977 年，是由数据库专业委员会举办的中国数据库领域的最高学术会议。自 1999 年数据库专委会成立以来，数据库专委会继承中国数据库多年来形成的优良传统，致力于办好这一传统的数据库盛会，为中国大陆、香港、台湾、澳门和海外华裔数据库研究者、开发者和用户提供一个大中华数据库论坛，交流有关数据库研究与应用的成果和经验，探讨数据库领域所面临的关键性挑战问题和研究方向，至今已成功举办 41 届。

第 42 届中国数据库学术会议（NDBC 2025）将于 2025 年 8 月 1 日-8 月 3 日在吉林长春举行。本届大会由中国计算机学会主办，中国计算机学会数据库专业委员会和吉林大学承办。本届大会将主要关注数据库领域所面临的新挑战、新问题和新方向，着力反映我国数据库技术研究的最新进展，为高校、科研院所、科技企业的数据库研究、开发和应用相关人员搭建交流平台。大会期间将设立多个技术论坛，并继续开设“系统演示”主题，开设“研究生论文辅导”研讨班等。届时还将邀请国内外数据库领域著名专家到会作专题报告。我们诚征数据库及其应用领域的论文、专题讨论与系统演示报告等。

### 2.1 征文范围（包括但不限于以下领域）

|                  |            |               |
|------------------|------------|---------------|
| 事务管理             | 云原生数据库     | 人工智能与数据库      |
| 向量数据库            | HTAP数据库    | 数据隐私和安全       |
| 查询执行与优化          | 并行与分布数据库系统 | 区块链与数据库       |
| 数据库系统性能评测        | 图数据管理      | 信息检索与数据库      |
| 近似和非确定性数据库       | 大数据管理与分析   | NoSQL数据管理     |
| 数据流管理            | 数据科学与工程    | 社交网络与推荐系统     |
| 新硬件下的数据管理        | 特定领域数据库    | 数据挖掘与知识发现     |
| 嵌入式与移动数据库        | 多媒体与多模态数据库 | 时空数据管理        |
| 支持机器学习与人工智能的数据管理 | 多模态数据处理    | 数据治理          |
| 以人为中心的数据管理       | 数据驱动的应用    | 其它与数据库相关的研究问题 |

## 2.2 征文要求

- 1) 论文应是未发表的研究成果，应包括中英文题目、中英文摘要、关键词、正文和参考文献。由于论文采用匿名评审，故论文中不能包含任何作者相关信息（如姓名、单位、电子邮件、通信地址、资助项目等），正文中也应当避免显式的论文自引(如“我们之前的工作[xx]”等)。
- 2) 大会同时接收中文和英文投稿。中文论文用 Word 排版，论文篇幅一般不超过 A4 幅面 8 页，格式请参考《计算机科学与探索》的投稿要求，可见《计算机科学与探索》网站的论文写作模板(<http://fcst.ceaj.org/CN/column/column6.shtml>)；英文论文格式请参考《Data Science and Engineering》的投稿要求 (<https://www.springer.com/journal/41019/submission-guidelines>)，英文论文 Latex 模板详细情况参见：<https://www.springernature.com/gp/authors/campaigns/latex-author-support>)。
- 3) 会议论文均采用网上提交方式，不支持电子邮件提交论文的方式。
- 4) 投稿网址：<https://conf.ccf.org.cn/ndbc2025/paper>

## 2.3 系统演示

本次大会设立系统演示程序（Demonstration Program），欢迎大家踊跃投稿。详细信息请见会议网站系统演示征文通知栏。

## 2.4 论文出版

会议录用论文拟推荐到《计算机学报》（正刊）、《软件学报》（正刊）、《计算机研究与发展》（正刊）、《浙江大学学报（工学版）》（正刊）、《清华大学学报》（正刊）、《计算机科学与探索》（正刊）、《计算机科学》（正刊）、《计算机应用》（正刊）、《计算机工程与应用》（正刊）、《计算机工程》（正刊）、以及英文期刊《Data Science and Engineering》（ESCI 收录）和《International Journal of Software and Informatics》等出版。

## 2.5 重要日期

论文提交截止时间：2025年5月19日

论文录用通知时间：2025年6月30日

论文修改与正稿提交时间：2025年7月15日

## 2.6 联系方式

会议网址：<https://conf.ccf.org.cn/ndbc2025>

联系人：张小利

邮件：[ndbc2025@126.com](mailto:ndbc2025@126.com)

电话：15043046596

## 3. 系统演示征文通知

系统演示（Demonstration Program）主要关注数据管理的创新技术与实现技术，展现我国在数据库系统研究方面的成就。通过演示原型系统，作者能够用更为有趣、直观的交互方式介绍自己的创新性系统研究工作。同时，系统演示可以促进和扩大国内同行在数据库系统实现和新技术应用上的创新与应用。在此诚征数据管理新技术及其新兴应用领域的系统演示报告。

### 3.1 征集范围

包括且不限于NDBC 2025 大会征文主题中所列出的所有专题。为了充分利用这个平台来展示系统的机会，诚邀各类新技术原型系统和针对新领域的应用系统参与系统演示。

### 3.2 投稿要求

- 1) 系统演示报告可以是未发表的研究成果，特别是最新数据管理技术的进展和实现，且应包括中英文题目、作者信息、中英文摘要、关键词、正文和参考文献等；也可以是近年来在 SIGMOD、VLDB、ICDE 等数据管理领域国际会议当中已经发表的研究成果。
- 2) 演示报告不同于研究论文。在报告中需要描述拟演示系统的体系架构，关键技术，系统演示步骤和演示环境等。
- 3) 未发表的研究成果以中英文撰写均可，论文篇幅一般不超过 4 页，中文论文格式参考《计算机科学与探索》的投稿要求，见《计算机科学与探索》网站的论文写作模板(<http://fcst.ceaj.org/CN/column/column6.shtml>)；英文论文格式请参考《Data Science and Engineering》的投稿要求（<https://www.springer.com/journal/41019/submit>），英文论文 Latex 模板详细情况参见：<https://www.springernature.com/gp/authors/campaigns/latex-author-support>；违反长度限制论文可被拒绝评审。
- 4) 系统演示论文均采用网上提交方式，不支持其他提交方式。
- 5) 投稿网址： <https://conf.ccf.org.cn/ndbc2025/paper>（投稿时，未发表研究成果的“论文领域”请选择“系统演示论文 A-未发表研究成果”，已发表研究成果的“论文领域”请选择“系统演示论文 B-已发表研究成果”）。

### 3.3 系统演示

- 1) 录用的系统演示报告不进行正式出版，但作者需要根据大会日程安排在指定日前完成系统和演示方案，并进行会议注册。
- 2) 为最大限度实现技术交流，系统演示将根据最终录用的数量安排交流研讨。
- 3) 会议组织方将根据观众的投票评选最佳系统演示报告奖（Best Demo Award）。

### 3.4 重要日期

论文提交截止时间：2025年6月15日

论文录用通知时间：2025年6月30日

### 3.6 联系方式

会议网址：<https://conf.ccf.org.cn/ndbc2025>

联系人：徐辰，张小利

邮件：cxu@dase.ecnu.edu.cn, ndbc2025@126.com

电话：16621057963, 15043046596

## 4. 注册通知

### 4.1 注册费标准

参会人员收费标准如下：

会议门票

| 门票类型       | 参会者身份   | 3.1 ~ 7.6 | 7.7 ~ 8.3 |
|------------|---------|-----------|-----------|
| 线下参会(不含晚宴) | CCF专业会员 | ¥2000     | ¥2250     |
|            | CCF学生会员 | ¥1300     | ¥1460     |
|            | 非会员专业人员 | ¥2500     | ¥2800     |
|            | 非会员学生   | ¥1600     | ¥1800     |

附加活动

| 附加活动名称 | 参会者身份   | 3.1 ~ 7.6 | 7.7 ~ 8.3 |
|--------|---------|-----------|-----------|
| 晚宴     | 不区分会员资格 | ¥300      | ¥300      |

### 4.2 缴费方式

参会者可在页面侧边或底边点击“报名”，进入中国计算机学会CCF会议系统页面进行注册，可通过线上支付宝、微信、易支付、银行汇款等方式支付，前三种线上支付都支持对公转账。

### 4.3 CCF 关于会议注册费的退费规定

欢迎扫描下方二维码加入 CCF。



由 CCF 主办及 CCF 下属机构承办（或协办）CCF 主办的会议或活动，涉及退费者，按如下规定执行：

- 1) 由于 CCF 的原因需要退款的，全额退款；
- 2) 由于参会者重复缴费，申请退回其中一笔者，CCF 收取所需退款费用 30%的管理费，其余退回；
- 3) 由于不能参会，会议举行 30 日前申请退款：收取 30%管理费，其余退回；
- 4) 由于不能参会，会议举行 7 日前申请退款：收取 50%管理费，其余退回；
- 5) 由于不能参会，会议举行前 7 日内（含）申请退款：不予退款。

如在办理退款前已开具发票，纸质发票须先退回 CCF 后方可办理退款手续，如系电子发票，则由 CCF 作废。

此规定从 2020 年 1 月 1 日起执行。

中国计算机学会  
2019 年 11 月 28 日

退费流程：退费要先及时发申请到本会议官方邮箱 [ndbc2025@126.com](mailto:ndbc2025@126.com)，然后会后由 CCF 进行退费处理。

CCF 建议：注册费可以转 CCF 会员费，如果这种方式，CCF 将不收取管理费用，由 CCF 会员部延续参会人员（本人）会员资格。

## 5. 会议日程简表

会议地址：长春国际影都万达文华酒店（金色大厅会议中心）

会议时间：2025年8月1日，星期五

| 时间          | 议程   | 地点          |
|-------------|--|-------------|
| 09:00-22:00 | 会议注册                                       | 酒店大堂        |
| 12:15-12:30 | 午餐   |             |
| 13:30-16:30 | 研究生辅导报告<br>论坛主席：邵鳌侠（北京邮电大学）                | 百花厅 1 会议室   |
| 13:30-16:30 | 专委常委会                                      | 金狮厅会议室      |
| 16:30-18:30 | 专委会会议                                      | 金鸡厅会议室      |
| 16:30-18:30 | 全国计算机名词委数据库与大数据审定分委员会会议<br>论坛主席：范举（中国人民大学） | 金狮厅会议室      |
| 19:00-21:00 | 企业之夜<br>论坛主席：袁野（北京理工大学）                    | 星光厅 1+2 会议室 |

会议时间：2025年8月2日，星期六

| 时间          | 议程  | 地点          |
|-------------|---|-------------|
| 08:30-09:00 | 开幕式   | 星光厅 1+2 会议室 |
| 09:00-09:10 | 合影  |             |
| 09:10-10:00 | 大会报告一：数据智能体:自主数据处理的新范式<br>主讲人：李国良（清华大学）                         | 星光厅 1+2 会议室 |
| 10:00-10:20 | 茶歇  |             |
| 10:20-11:10 | 大会报告二：大模型:当前数据塑造智能的一条成功路径<br>主讲人：薛向阳（复旦大学）                      | 星光厅 1+2 会议室 |
| 11:10-12:00 | 大会报告三：从应用视角看数据库发展与未来<br>主讲人：潘安群（腾讯云数据库）                         | 星光厅 1+2 会议室 |
| 12:00-14:00 | 午餐  |             |
|             | 分论坛： 第五生产要素背景下的数据技术论坛<br>论坛主席：钱卫宁（华东师范大学）<br>周烜（华东师范大学）         | 金鸡厅 1 会议室   |
|             | 分论坛： 大规模图数据管理与分析论坛<br>论坛主席：赵翔（国防科技大学）<br>张文涛（北京大学）              | 金鸡厅 2 会议室   |
|             | 分论坛： 时序数据智能管理与分析论坛<br>论坛主席：胡吉林（华东师范大学）<br>宋轩（吉林大学）              | 百花厅 1 会议室   |
| 14:00-17:30 | 分论坛： 数据库系统新技术论坛<br>论坛主席：柴云鹏（中国人民大学）<br>卢卫（中国人民大学）               | 百花厅 2 会议室   |
|             | 分论坛： CCF-华为胡杨林基金数据库专项论坛<br>论坛主席：张岩峰（东北大学）<br>李士福（华为技术有限公司）      | 金熊厅会议室      |
|             | 分论坛： CCF-阿里云瑶池科研基金论坛<br>论坛主席：杜小勇（中国人民大学）<br>李飞飞（阿里云）            | 金狮厅会议室      |
|             | 分论坛： CCF-蚂蚁科研基金数据库专项论坛<br>论坛主席：周傲英（华东师范大学）<br>杨传辉（北京奥星贝斯科技有限公司） | 金棕榈厅会议室     |
|             | 分论坛： 论文报告一  | 奥斯卡厅会议室     |
| 15:10-15:30 | 茶歇  |             |
| 18:00       | 晚宴  | 星光厅 1+2 会议室 |

会议时间：2025年8月3日，星期日

| 时间          | 议程   | 地点          |
|-------------|--|-------------|
| 08:30-12:00 | 分论坛： 软件学报专刊论坛<br>论坛主席：李国良（清华大学）<br>张蓉（华东师范大学）  | 金鸡厅 1 会议室   |
|             | 分论坛： 系统演示<br>演示主席：徐辰（华东师范大学）<br>赵翔（国防科技大学）   | 长廊          |
|             | 分论坛： 萨师煊优秀学生论文汇报   | 金鸡厅 2 会议室   |
|             | 分论坛： 论文报告二   | 百花厅 1 会议室   |
|             | 分论坛： 教育论坛<br>论坛主席：王鑫（天津大学）<br>张超（中国人民大学）   | 百花厅 2 会议室   |
|             | 分论坛： 贝叶斯数据分析论坛<br>论坛主席：岳昆（云南大学）<br>俞奎（合肥工业大学）  | 金熊厅会议室      |
|             | 分论坛： 新一代数据库与数据赋能技术国际论坛<br>论坛主席：于明鹤（东北大学）<br>卢卫（中国人民大学）                                   | 金狮厅会议室      |
|             | 分论坛： 数据库开源技术论坛<br>论坛主席：王晨（清华大学）<br>彭煜玮（武汉大学）   | 金棕榈厅会议室     |
|             | 分论坛： 移动云大云海山数据库合作生态论坛<br>论坛主席：伍赛（浙江大学）<br>丁岩（中移（苏州）软件技术有限公司）                             | 奥斯卡厅会议室     |
| 10:10-10:30 | 茶歇   |             |
| 12:00-13:30 | 午餐   |             |
| 13:30-14:20 | 大会报告四：Next-Generation Blockchains: Trends, Challenges, and Solutions<br>主讲人：黄铭钧（新加坡国立大学） | 星光厅 1+2 会议室 |
| 14:20-15:10 | 大会报告五：Database Technologies for Healthcare<br>主讲人：横田治夫（东京工业大学）                           | 星光厅 1+2 会议室 |
| 15:10-15:30 | 茶歇   |             |
| 15:30-17:00 | 颁奖仪式与闭幕式   | 星光厅 1+2 会议室 |

## ■ 6. 研究生辅导报告

### 6.1 论坛日程

地点：百花厅 1 会议室

时间：2025 年 8 月 1 日，星期五，下午 13:30 ~ 16:30

论坛主席：邵鳌侠（北京邮电大学）

| 时间          | 报告题目                     | 报告人 | 单位     |
|-------------|--------------------------|-----|--------|
| 13:30-13:40 | 论坛主席致辞                   | 邵鳌侠 | 北京邮电大学 |
| 13:40-14:10 | 问题驱动的科学的研究               | 张伟  | 华东师范大学 |
| 14:10-14:40 | 研究生做科研的几点建议              | 柯翔宇 | 浙江大学   |
| 14:40-15:00 | 茶歇                       |     |        |
| 15:00-15:30 | 论文撰写与科学的研究初探             | 张奥千 | 北京理工大学 |
| 15:30-16:00 | 研究生科研入门<br>——第一篇顶会论文怎么写？ | 刘钰  | 北京交通大学 |
| 16:00-16:30 | 研究生科研：如何找论文与读论文          | 张心怡 | 中国人民大学 |

### 6.2 论坛简介

本论坛拟从论文写作、科研任务构思、学术研究方法等方面出发，介绍研究生科研方法，为研究生同学提供学术辅导。

## 6.3 论坛主席

邵鳌侠



**专家介绍:** 邵鳌侠, 北京邮电大学计算机学院教授, 国家级青年人才, 博士生导师。研究方向为图数据管理、大规模图计算、AI4Sys 等。主持 3 项国家自然科学基金项目、1 项科技创新 2030“新一代人工智能”重大项目子课题以及多项 CCF/CAAI 产学研合作基金。在数据库和人工智能领域的著名会议和期刊已发表高质量学术论文 100 余篇, 其中第一作者/通讯作者发表 CCF-A 类期刊和会议论文 30 余篇, 出版英文学术专著 1 部; 获数据库领域顶级国际会议 VLDB 2022 最佳常规研究论文奖、数据库领域重要国际会议 DASFAA 2020 最佳学生论文奖、CCF NDBC 2016 萨师煊优秀学生论文奖等国内外学术会议论文奖 7 项。入选了北京市科技新星、小米青年学者, 获中国电子学会科学技术奖科技进步奖一等奖、北京市高等教育教学成果奖二等奖、ACM SIGMOD 中国新星奖等。指导学生获全国大学生计算机系统能力大赛- (首届) 数据库管理系统设计赛一等奖、第四届实时计算 Flink 挑战赛亚军; 长期担任多个重要国际会议的程序委员和著名国际期刊的审稿人。

## 6.4 论坛报告

### 问题驱动的科学研究

张伟



**报告摘要：**科学研究紧密围绕着问题展开，问题本身的价值对于研究价值具有重要影响。本次分享将结合组内近年来的研究与应用工作，分享在问题驱动的科学研究所的一些个人体会，主要涉及两个大的方面：首先在问题寻找与评价方面，将探讨如何寻找问题，如何评判问题的研究价值和潜在影响力；其次在研究案例方面，将介绍团队从应用问题驱动和理论问题驱动两个角度，如何进行真实问题驱动的科学研究所工作。

**专家介绍：**张伟，华东师范大学计算机学院教授、博士生导师，CCF 高级会员，数据库专委会执行委员。博士毕业于清华大学计算机科学与技术系，随后入选华东师范大学“双百人才计划”。目前主要研究方向包括：用户大数据挖掘、可解释与高效机器学习、大语言模型等。近年先后主持国家重点研发计划课题、国家自然科学基金重大研究计划培育项目等。共发表学术论文百余篇，其中第一/通讯作者发表 CCF A 类国际期刊和会议论文 50 余篇。担任 NeurIPS、ACL 等 CCF-A 类国际会议程序委员会领域主席，指导学生获得多项国际国内学术竞赛奖项。

## 研究生做科研的几点建议

柯翔宇



**报告摘要：**本报告旨在为研究生同学提供一套系统、实用的科研入门与进阶指南。

报告将首先聚焦科研的核心方法论，涵盖关键科研环节的精髓：从如何高效进行文献调研与批判性阅读，到科学问题的凝练与创新性研究选题的发掘；从严谨实验设计（或理论建模）与数据获取，到结果分析与验证的有效策略。报告将特别强调批判性思维和系统性方法在科研过程中的核心地位。其次，报告将深入探讨高质量学术论文的写作之道。内容将贯穿论文写作全周期：包括论文结构的精准把握、各章节的写作要点与技巧、图表制作的规范与艺术、学术语言的精炼与表达，以及有效应对审稿意见与修改的策略。目标是帮助研究生掌握将复杂研究成果清晰、准确、有说服力地呈现给学术共同体的能力。此外，报告将紧密结合报告人自身的研究领域——数据管理与数据智能的交叉融合——进行实例化探讨。本报告期望通过方法论的梳理、写作技巧的传授，以及结合前沿研究方向的生动案例剖析，助力研究生同学们建立坚实的科研基础，提升科研效率与论文质量，更加自信地开启并推进自己的学术探索之旅。

**专家介绍：**柯翔宇，浙江大学计算机科学与技术学院、软件学院平台“百人计划”研究员，博士生导师，隶属于浙江省大数据智能计算全省重点实验室。申请人博士毕业于新加坡南洋理工大学，先后于新加坡南洋理工大学和新加坡国立大学担任博士后，长期从事大图数据智能管理与分析领域的研究工作，任中国计算机学会（CCF）数据库专委、信息系统专委执行委员，大数据专委通讯委员。柯翔宇博士在数据库、数据挖掘领域顶级国际会议和期刊上发表论文三十余篇；获宁波市（副省级）甬江人才工程科技创新领域青年创新人才项目支持，任宁波市科技项目评审专家，宁波市拔尖人才，获CCF-阿里云瑶池科研基金支持。

## 论文撰写与科学初探

张奥千



**报告摘要：**本次报告，讲者将结合自己博士生、博士后再到研究生导师的经历，与大家分享在数据库领域的研究经验与思考。为研究生同学在论文撰写、科研生活两个方面提供建议和参考。首先，讲者将讨论数据库领域学术论文的写作技巧与经验，其次，将结合国内外科研生活的经历，分享自己对于研究生阶段科研生活的看法和建议。最后，讲者期待与大家进行开放式交流，共同进步、共同提高。

**专家介绍：**张奥千，北京理工大学计算机学院副研究员，于 2018 年从清华大学软件学院获得博士学位，随后于加拿大滑铁卢大学从事博士后研究，2020 年入职北京理工大学，主要从事数据库、数据治理与数据挖掘方向的研究。在计算机国际顶级会议和期刊（CCF-A，包括 SIGMOD、VLDB、ICDE、VLDBJ、TKDE、TODS）上共发表论文十余篇，并且在多个计算机顶级会议和期刊上担任程序委员和评审工作。主持国家自然科学基金青年项目 1 项，国家重点研发计划“区块链专项”课题 1 项，CCF-华为胡杨林基金数据库专项 1 项。2024 年获 ACM 北京分会新星奖。

## 研究生科研入门——第一篇顶会论文怎么写？

刘钰



**报告摘要：**硕士和博士研究生期间的第一篇论文尤其是第一篇顶会论文的发表，对研究生科研生涯的入门起步至关重要。本报告分别从论文的技术创新和文章撰写两个角度分享一些经验和启发。从技术创新角度，给出论文贡献点、算法模型系统设计、细节优化和代码实现等层面的一些建议；从文章撰写角度，从论文整体组织、章节段落逻辑、亮点突出和升华等层面分享一些经验。希望能帮助研究生们快速入手第一篇高质量论文。

**专家介绍：**刘钰，博士，现任北京交通大学计算机科学与技术学院副教授、博士生导师，研究方向为大规模图计算与学习关键技术。2018 年博士毕业于中国人民大学信息学院，于 2018-2021 年任北京大学博雅博士后，兼聘为北京大学重庆大数据研究院研究员。在图相关方向发表国际顶级会议和期刊论文（CCF A 类）多篇，包括 VLDB、SIGMOD、TKDE、ICDE、VLDB Journal 等。主持国家级科技项目、国家自然科学青年基金、CCF-腾讯犀牛鸟科研基金、腾讯犀牛鸟专项研究计划、北京大学医信交叉种子基金等研究项目，作为课题骨干参与多项国家自然科学重点基金和北京通用人工智能研究院、华为 2012、MSRA Research Grant 等多个高水平研究项目。参与的华为合作研究获 2023 年（第 3 届）中国工业与应用数学学会（CSIAM）应用数学成果落地认证，主持的 CCF-腾讯犀牛鸟科研基金获评优秀专利奖。参与编著《大规模图数据管理与分析》教材。

## 研究生科研：如何找论文与读论文

张心怡



**报告摘要：**本报告聚焦研究生在科研初期面临的关键任务—科研文献的搜集与阅读。首先，介绍多种高效的文献检索策略；其次，讲解如何判断文献质量、建立阅读优先级，并有效拓展知识广度；最后，结合“三遍阅读法”等实用技巧，分享科研论文的高效阅读与信息提取方法。通过本报告，学生将系统掌握从“找资料”到“用资料”的完整流程，切实提升信息获取与知识整合能力，为后续科研打下坚实基础。

**专家介绍：**张心怡，中国人民大学信息学院讲师，吴玉章青年英才，CCF 数据库专委委员，长期从事数据库系统、AI4DB（人工智能赋能的数据库系统）方向的研究，在该领域发表 CCF A 类论文十余篇，入选 2025 年“高创计划”青年人才托举工程、2025 年中国计算机学会数据库专委会博士学位论文激励计划、2023 年字节跳动奖（全球仅 13 人）、2022 年微软学者提名（亚太地区仅 21 人）、北京大学优秀博士论文、NDBC 最佳学生论文奖。主持 CCF-阿里云瑶池基金、CCF-华为胡杨林基金，担任 VLDB、TKDE、TODS、JCST 等会议及期刊的审稿人，并担任 ICDE、WWW 程序委员会成员。

## ■ 7. 企业之夜

### 7.1 论坛日程

地点：星光厅 1+2 会议室

时间：2025 年 8 月 1 日，星期五，下午 19:00 ~ 21:00

论坛主持：袁野（北京理工大学）、秦建斌（深圳大学）、许建秋（南京航空航天大学）

论坛主席：袁野（北京理工大学）

宴会形式：自助晚餐

参会人员：NDBC 2025 参会者免费入场

| 时间          | 议程                  | 人员      |
|-------------|---------------------|---------|
| 19:00-19:05 | 主持                  | 秦建斌、许建秋 |
| 19:05-19:10 | 主席致辞                | 袁野      |
| 19:10-19:20 | 签约仪式                | 赞助商签约   |
| 19:20-19:50 | 企业分享、互动抽奖（视频、发言、抽奖） |         |
| 19:50-21:00 | 自由交流                |         |
| 21:00       | 晚宴结束                |         |

## ■ 8. 大会报告

### 8.1 日程安排

地点：星光厅 1+2 会议室

时间：2025 年 8 月 2 日 上午 09:10-12:00

| 时间          | 报告题目                | 报告人 | 单位     |
|-------------|---------------------|-----|--------|
| 09:10-10:00 | 数据智能体：自主数据处理的新范式    | 李国良 | 清华大学   |
| 10:00-10:20 | 茶歇                  |     |        |
| 10:20-11:10 | 大模型：当前数据塑造智能的一条成功路径 | 薛向阳 | 复旦大学   |
| 11:10-12:00 | 从应用视角看数据库发展与未来      | 潘安群 | 腾讯云数据库 |

时间：2025 年 8 月 3 日 下午 13:30-15:10

| 时间          | 报告题目   | 报告人  | 单位      |
|-------------|--|------|---------|
| 13:30-14:20 | Next-Generation Blockchains: Trends, Challenges, and Solutions | 黄铭钧  | 新加坡国立大学 |
| 14:20-15:10 | Database Technologies for Healthcare                           | 横田治夫 | 东京工业大学  |

## 8.2 大会论坛报告

### 数据智能体：自主数据处理的新范式

李国良



**报告摘要：**传统数据处理系统存在三大核心局限：封闭世界查询假设、受限的数据支持和被动式管理模式，导致其难以满足多样化的业务需求、受限的数据处理范畴以及僵化的管理机制。人工智能技术的发展为解决这些问题提供了新契机，其语义理解、推理、编排与生成等能力，为构建支持开放世界数据管理、多模态数据处理和自驱动范式的新一代数据管理系统奠定了基础。为此，我们提出数据智能体（Data Agent），通过整合语义理解、智能推理与编排、智能生成、多智能体协同等能力自主完成各类数据处理任务。本报告将详细介绍数据智能体的关键技术与系统构建方法。

**专家介绍：** 李国良，清华大学计算机系教授，副主任，CCF 数据库专委副主任，ACM Fellow, IEEE Fellow。主要研究方向为数据库。在顶级会议和期刊上发表论文 200 余篇，他引 20000 余次，入选爱思唯尔 2014-2024 年中国高被引学者榜单。曾主持国家杰出青年基金(延续资助)、重点研发、青年 973、自然基金重点等项目。获得了 VLDB 2017 Early Career Research Contributions Award (VLDB 青年贡献奖，亚洲首位获奖者)、IEEE TCDE Early Career Award (IEEE 数据工程领域杰出新人奖，亚洲首位获奖者)。担任 VLDB Journal、IEEE TKDE 等编委，SIGMOD 2021 大会主席，ICDE 2027 程序委员会主席。获得过数据库领域重要国际会议 15 项论文奖项，如 SIGMOD 2024 研究亮点奖、ICDE 2025 最佳论文亚军、VLDB 2023 工业最佳论文亚军、CIKM 2017 最佳论文等。

## 大模型：当前数据塑造智能的一条成功路径

薛向阳



**报告摘要：**今天，以大语言模型为代表的人工智能新技术，其应用迅速渗透到科研、教育、医疗、制造、金融等众多领域，成为推动人类社会发展的新质生产力，深刻重塑我们的生产生活方式。本报告将简要介绍大语言模型的基本原理，剖析数据在模型训练和推理过程中的作用，分析数据规模与模型性能之间的缩放规律，同时探讨单纯依赖数据驱动的大模型的能力边界，并在此基础上，进一步思考未来具有“人类水平”的人工智能的可能发展路径。

**专家介绍：**复旦大学计算与智能创新学院教授，CCF 杰出会员、CSIG 常务理事、上海市人工智能学会副理事长。现任复旦大学大数据研究院、类脑智能科学与技术研究院副院长。主要研究兴趣包括计算机视觉、多模态大模型、具身智能等。

## 从应用视角看数据库发展与未来

潘安群



**报告摘要：**从早期企业服务时代，历经 PC 互联网时代与移动互联网时代，现如今已然进入 AI 时代，浪潮一波接着一波，深刻地改变着我们的生产和生活方式。与此同步，承载着海量数据、支撑着各类应用运行的数据库技术，也始终在推陈出新，不断迭代其产品形态和能力，以适应日新月异的应用和业务发展需求。本报告将深入聚焦于应用视角，与您共同探讨在当前以及未来的 AI 应用环境下，数据库需要满足哪些核心要求。更重要的是，我们将剖析在 AI 浪潮中，我们应当如何构建一个具备前瞻性、高性能且能够充分赋能 AI 应用的数据库体系，以应对数据爆炸式增长和复杂计算带来的挑战。

**专家介绍：**腾讯云数据库研发负责人，曾获深圳市科技进步一等奖、湖北省科技进步奖一等奖，中国专利金奖，广东专利金奖，负责腾讯云数据库 TDSQL 技术研发，拥有超过 18 年分布式计算和分布式数据库研发经验。他带领团队打造的自主可控分布式数据库 TDSQL，分别以 8.14 亿 tpmC 和 7260 万 QphDs 的性能刷新 TPC-C 和 TPC-DS 测试世界纪录，是业内首个应用于互联网银行核心交易系统、首个进入银行传统核心系统的分布式数据库，首个帮助国有大行进行数据库国产化替代的产品，目前已经为超过 30000 家金融政企客户提供数据库服务，助力行业数据库实现全面自主可控。

## Next-Generation Blockchains: Trends, Challenges, and Solutions

黄铭钧



**报告摘要：**Blockchain technology has rapidly evolved towards next-generation systems, yet realizing their full potential requires overcoming fundamental challenges in scalability, security, and interoperability. In this talk, I will outline these critical hurdles and showcase key research contributions aimed at providing robust solutions. I will also present novel approaches to enhance system performance, protect smart contract and transaction security, and enable efficient cross-chain communication, paving the way for a more advanced and trustworthy decentralized ecosystem. I will also talk briefly on my move to NingBo.

**专家介绍：**黄铭钧，中国科学院外籍院士、新加坡科学院院士、新加坡工程院院士、欧洲科学院外籍院士、ACM Fellow、IEEE Fellow、CCF 会士。澳大利亚莫纳什大学计算机科学专业博士毕业后，曾任新加坡国立大学李光前百年纪念讲席教授、浙江大学长江讲座教授，现任浙江大学求是讲席教授、浙江大学软件学院教授、清华大学杰出访问教授、北京大学讲座教授等职。本人长期从事数据库性能问题、索引技术、内存数据管理、人工智能、云计算和并行分布式系的研究与在智慧医疗、金融数据分析等领域的高级应用。2014 年至今，耗时多年研发了新一代分机器学习平台 Apache SINGA，助力人工智能在互联网数据分析中的实践；2016 年，发布了世界首个区块链评测系统 BlockBench，并在 GitHub 开源代码，引起学术界和工业界的广泛关注。在顶级会议和杂志上发表 300 多篇论文，被引用 33232 次，H 指数 96。作为唯一的华人和亚太地区数据库专家在 2008、2013、2018、2023 年连续四届参加了国际数据库顶级专家闭门会。获得国际计算机学会（ACM）数据管理专业委员会（SIGMOD）颁发的贡献奖（2009 年）、Edgar F. Codd 创新奖（2020 年）、系统奖（2024 年）三大数据库领域最高荣誉。同时，本人也曾获新加坡总统科学奖（2011 年），IEEE 计算机学会 Tsutomu Kanai 奖（2012 年），IEEE TCDE CSEE (Computer Science, Engineering, and Education) 影响奖（2014 年），中国计算机学会海外杰出贡献奖 (CCF Outstanding Contributions Award) (2016 年)。

个人主页：<https://ooibc88.github.io/>

## Database Technologies for Healthcare

横田治夫



**报告摘要:** The impact that large language models (LLMs) have had on society is profound. Today, a wide range of tasks across various domains are being supported by systems powered by LLMs. In the field of healthcare, determining how best to leverage these models has become a major area of exploration and innovation.

At the same time, LLMs are often perceived as "black boxes" due to the opacity of their internal workings. This poses challenges when attempting to directly apply their outputs to sensitive domains such as medicine, where reliability, interpretability, and accountability are critical.

In this context, a variety of techniques developed over the years in the field of databases can offer valuable support in medical applications. Traditional methods such as OLAP (Online Analytical Processing), data cube analysis, association rule mining, sequential pattern mining, variant analysis, outlier detection, and matrix factorization have long been used to extract insights from structured data.

By examining how such data-centric techniques can contribute to healthcare support, we can also explore their potential complementarity with LLM-based approaches. For example, structured pattern mining may provide interpretable evidence to validate or supplement human and LLM-generated suggestions, while anomaly detection and variant analysis may help identify edge cases where LLMs struggle. This interplay between statistical analysis and deep learning models could offer a more robust and trustworthy framework for advancing medical informatics.

**专家介绍:** Haruo Yokota received his B.E., M.E., and Dr.Eng. degrees from the Tokyo Institute of Technology in 1980, 1982, and 1991, respectively. He joined Fujitsu Ltd. in 1982 and served as a researcher at ICOT for the Japanese Fifth Generation Computer Project from 1982 to 1986, and at Fujitsu Laboratories Ltd. from 1986 to 1992.

From 1992 to 1998, he was an Associate Professor at the Japan Advanced Institute of Science and Technology (JAIST). He moved to the Tokyo Institute of Technology in 1998, where he became a Full Professor in the Department of Computer Science. He served as Dean of the School of Computing at the Tokyo Institute of Technology from 2018 to 2022 and retired in 2023.

He is currently an Emeritus Professor at the Tokyo Institute of Technology and a Specially Appointed Professor at Josai University. His research interests include data engineering, information storage systems, and dependable computing.

He has held various leadership roles, including Chair of the ACM SIGMOD Japan Chapter, Trustee Board Member of the Information Processing Society of Japan (IPSJ), Editor-in-Chief of the Journal of Information Processing, and Associate Editor of the VLDB Journal. He is currently Chair of the Database Society of Japan (DBSJ), a Fellow of both IEICE and IPSJ, a Senior Member of IEEE, and a member of IFIP WG10.4, JSAI, ACM, and ACM SIGMOD.

## ■ 9. 第五生产要素背景下的数据技术论坛

### 9.1 论坛日程

地点：金鸡厅 1 会议室

时间：2025 年 8 月 2 日，星期六，下午 14:00 ~ 17:00

论坛主席：钱卫宁（华东师范大学）、周烜（华东师范大学）

| 时间          | 报告题目                    | 报告人       | 单位       |
|-------------|-------------------------|-----------|----------|
| 14:00-14:10 | 论坛主席致辞                  | 钱卫宁<br>周烜 | 华东师范大学   |
| 14:10-14:40 | 对数据及其价值的再认识与再实现         | 周傲英       | 华东师范大学   |
| 14:40-15:10 | 协同智能与通证经济：概述与前沿         | 朱飞达       | 新加坡管理大学  |
| 15:10-15:30 | 茶歇                      |           |          |
| 15:30-16:00 | 复杂系统数据治理的范式演进与探索        | 贾晓丰       | 北京市大数据中心 |
| 16:00-16:30 | 一汽数字化转型中的数据技术           | 徐智        | 中国一汽     |
| 16:30-17:00 | AI智能体的法律治理路径：以最小必要原则为中心 | 赵精武       | 北京航空航天大学 |

### 9.2 论坛简介

2020 年 4 月 9 日中共中央、国务院公布《关于构建更加完善的要素市场化配置体制机制的意见》。作为我国最高级别的文件，《意见》把数据定义为和劳动力、土地、资本一样的生产要素。作为生产要素的数据已经超越了技术的范畴，在这样的背景下讨论数据相关的技术需要突破传统的关于数据的认知。数据库自诞生以来一直被认为是管理数据的基础设施，数据作为生产要素为数据库带来挑战和机遇。常规的关于数据的理解是“以数为据”，数据作为依据，这与数据作为生产要素有很大的差异。作为依据，数据是副产品，越少越好；作为生产要素，数据是越多越好。早期的 OLTP 数据库中存储的就是作为副产品的依据，数据库取得成功之后发展出的数据挖掘、商务智能从某种意义上而言，可以认为是作为副产品的数据的“无意识的变废为宝”。五年多前公布的《意见》和后来成立的国家数据局，根本意义就在于实现数据“有意识的价值释放”。要做到这一点，首先就必须突破传统的从技术层面对数据的认知，要上升到经济学和哲学层面来重新认识数据。生产要素的本质就是创造价值和参与分配，这就不仅仅是技术问题，必然涉及到权属、权利等根本问题。回顾人们对数据认识的心路历程，互联网起了重要的作用，是互联网彰显了数据的重要性，大数据是说数据很重要。数据的重要性堪比引发第二次工业革命的电力，有了电，蒸汽时代所做的事情都可以重做一遍，还产生了 IT 这个在蒸汽时代不可能有的行业。

审视当下的人工智能热潮，我们似乎又处在交流电刚被发明的那个时代，未来已来，一切重来。本次论坛，我们邀请了来自学术界、产业界和政策部门的专家，从不同的视角来解读在数据作为生产要素的背景下数据技术所扮演的角色和发挥的作用，为我国数据库学者和研发人员提供有益的启示。

### 9.3 论坛主席

钱卫宁



**专家介绍：**钱卫宁，华东师范大学教授、博士生导师，数据科学与工程学院院长。在复旦大学获得计算机科学与技术学士、硕士与博士学位。2020 年入选上海市优秀学术带头人，2021 年入选国家级高层次人才计划。作为项目负责人主持包括国家重点研发计划项目、国家自然科学基金项目重点项目在内的多项科研项目。目前担任中国计算机学会数据库专委会常务委员，中国系统工程学会数据科学与知识系统工程专委会副主任委员，中国计量测试学会计算基准专委会副主任委员，教育部人工智能科技创新专家工作组成员。曾获国家科技进步二等奖 1 次（第二完成人）、上海市科技进步一等奖 2 次（第一、九完成人），以及教育部科技进步一等奖、教育部技术发明一等奖等奖项。研究兴趣包括可扩展事务处理，大数据管理系统基准评测，海量数据分析处理及其应用，以及数据驱动的计算教育学。

周烜



**专家介绍：**周烜，华东师范大学教授、数据科学与工程学院副院长。2005 年于新加坡国立大学取得博士学位，先后在德国 L3S 研究中心、澳大利亚联邦科工组织（CSIRO）和中国人民大学信息学院任职，2017 年入职华东师范大学。长期专注于数据库系统领域的研究，曾负责多个国内外的科研项目和工业合作项目，开发的数据管理系统在多个行业落地应用，大量科研成果发表于国际主流学术期刊和会议，获 2019 年国家科技进步二等奖、2023 年国家级教学成果二等奖，入选国家级高层次人才计划。

## 9.4 论坛报告

### 对数据及其价值的再认识与再实现

周傲英



**报告摘要：**数据是第五生产要素，这一论断彰显了我们对于数据的认识达到一个新的高度。与作为第三生产要素的资本相比，我们对数据的认识还处于初级阶段。回顾历史，二十多年来互联网的发展使我们认识到数据的重要性。数据之于数字化，正如电力之于电气化，数据是人类文明史上位列汽力、电力之后的第三个重要的动力，将把人类带入数字文明时代。重新认识数据才能有效推进数字化转型，只有在认知层面把数据从“副产品”提升到“新要素”，才能把无意识的“变废为宝”变成有意识的“价值释放”。对数据的认知突破是释放数据作为生产要素的威力的前提，数据作为新的能源动力带来新质生产力，根本标志是生产力的数量级提升。

**专家介绍：**周傲英，华东师范大学数据学院教授，中国计算机学会会士、常务理事、数据库专委会主任。现担任第八届国务院学科评议组成员、第八届教育部科技委委员、上海市计算机学会理事长、上海市人工智能与社会发展研究会会长、《计算机学报》副主编。曾入选教育部长江学者特聘教授，获得过国家自然科学基金委员会杰出青年基金。曾任复旦大学计算机系主任（1999-2002）、华东师范大学副校长（2016-2023）。主要研究兴趣包括数据库、数据管理、区块链、数字化转型、金融科技、教育科技等。获得过国家科技进步二等奖、国家教学成果二等奖及多项教育部和上海市技术发明或科技进步一等奖。

## 协同智能与通证经济：概述与前沿

朱飞达



**报告摘要：**人工智能带来了诸多令人惊叹的创新的同时也引发了关于治理问题的广泛关注，尤其是在其应用场景日益趋向协作化的背景下。当数据、模型和计算资源来自不同主体时，如何同时提供信任与激励机制？本次演讲将概述人工智能与 Web3 在协同智能与通证经济背景下的潜在融合，介绍其中的挑战、核心技术组件以及最前沿的研究进展。我们将重点介绍在去中心化联邦学习场景下的数据协作新协议，这些协议兼顾隐私感知的数据验证机制与基于代币的激励设计。同时，作为“信任内置”数据与人工智能协作的重要基础，分布式账本技术所涉及的安全性问题也将被探讨，并介绍相关的链上数据分析新算法。此外，我们还将通过若干应用案例，展示协同智能与通证经济之间的深度融合实践。

**专家介绍：**朱飞达博士目前是新加坡管理大学计算与信息系统学院的 Aptos Move 讲席教授与副院长，也是新加坡科技研究局 A\*Star 兼职高级首席科学家。主要研究方向包括人工智能和机器学习技术与治理，Web3 与可持续数智经济，数据资产和数据要素，区块链和分布式账本技术，社交网络及新媒体数据挖掘与分析，数据驱动的商业智能和创新研究。至今已在数据库和数据挖掘国际期刊和国际会议上发表学术论文 100 余篇，并获得 PAKDD’19 Early Career Award, DASFAA’16, ICDE’07 和 PAKDD’07 的多个最佳论文奖。曾担任 100 多个国际会议的资深程序委员会委员和多个国际会议的大会主席，包括 DASFAA’25 的 PC Chair, ACM SIGKDD’21 和 IEEE ICDM’18 的 General Chair。博士毕业于美国伊利诺伊大学香槟分校 University of Illinois at Urbana-Champaign (UIUC)，师从韩家炜 (Jiawei Han) 教授。

## 复杂系统数据治理的范式演进与探索

贾晓丰



**报告摘要：**数据作为新型生产要素，其基础性、系统性和活跃性对传统治理体系提出了挑战，尤其面对数源巨量分散、数据异质敏感、关系隐性多变等的超大城市复杂系统，需要新的技术框架来解决数据与场景之间的感知、连接、控制、计算和保护问题。本报告分析数据跨域流通的形态、路径和语义，介绍数据治理的范式演进，阐述以数据共生为特征的数据治理 5.0 框架。

**专家介绍：**贾晓丰，教授级高工，博导，北京市大数据中心数据治理部部长，大数据协同安全技术国家工程实验室首席研究员，CCF 数据治理与发展技术委员会常委，CAA 理事、联邦智能专委会副主任。主要研究方向聚焦复杂系统数据治理、数据驱动的人工智能，主持国家自然科学基金、国家重点研发计划和北京市重点项目/课题 20 余项，以第一完成人获省部级自然科学一等奖、科技进步一等奖，被授予全国先进个人（中共中央 国务院 中央军委）、全国先进集体负责人（团中央）。

## 一汽数字化转型中的数据技术

徐智



**报告摘要：**数字化转型是信息技术与业务交叉的演进过程，数据资源是被关注对象。但是，传统用于管理数据资源的数据库技术存在自身的局限性，已经难以跟上数据在规模和多样性上的发展速度。在一汽集团数字化转型的实践中，总结了一些数据技术的经验与规律，期待与业内专家共享和研讨，主要内容如下：

- 汽车行业市场数据缝合方法——数据的时间价值；
- 一汽集团报表体系——统计范围与时效的关系；
- 一汽集团实时数据湖架构——如何缩短计算时间；
- 基于企业数据的领域大模型——数据压缩诞生人工智能；
- 一汽集团数据技术的代际规律——数据技术的投资周期。

**专家介绍：**徐智，中国第一汽车集团有限公司-体系数字化部数据科学工程师，在读工程博士，专研数据技术十数年，洞察数据的基本规律，对数据可观测性、架构可行性、业务可接受度都有实践创新和成功经验。精通数据技术理论，在数据湖仓、数据中台、大数据平台、数据库私有云领域，设计和实施方案都处于业界领先水平。

## AI智能体的法律治理路径：以最小必要原则为中心

赵精武



**报告摘要：**AI 等信息技术的创新发展正在持续对既有的个人信息保护制度提出挑战，最小必要原则在 AI 智能体领域的适用难题则是其中的典型代表。AI 智能体在提升 AI 信息服务智能化、自动化程度的同时，却也带来个人信息保护的新问题：AI 智能体的服务模式属于“个性化服务”，故而需要收集和处理尽可能多的个人信息。但是，由于 AI 智能体基础功能和个人信息处理目的的不确定性，以“确属处理个人信息处理目的所必要”为判断逻辑的最小必要原则难以适用。在该场景下，最小必要原则并未因与产业实践脱节而失灵，而是既有的解释论侧重强调“最小”和“必要”的具体要求，忽视了该原则所指向的个人信息保护目标。所谓的“最小”不是指数量层面的最小，而是强调在多种技术方案中选择“对个人权益影响最小”且收集个人信息数量相对最小的方案。所谓的“必要”并不是仅以处理目的作为判断标准，而是结合 AI 智能体应用场景、信息服务合同、技术保护措施等要素进行综合判断，需要结合技术发展考量究竟如何在实践层面实现“保护个人信息”和“促进个人信息合理利用”这两项立法目标。

**专家介绍：**赵精武，北京航空航天大学法学院副教授、院长助理。中国法学会网络与信息法学研究会理事兼副秘书长、工信部重点实验室副主任，北京航空航天大学网络空间国际治理研究基地副主任，北京科技创新中心研究基地副主任。学术研究方向为数字法学。入选全国广播电视台和网络视听行业青年创新人才、北京市高层次人才、北京市社会科学基金青年学术带头人、北京市法学会“百名法学青年英才”培养计划、北京市国家治理青年人才培养计划、北航青年五四奖章。译有《所有权的终结：数字时代的财产保护》《数字正义：当纠纷解决遇上互联网》等，在《中国法学》《中外法学》《法律科学》《现代法学》《法制与社会发展》《法学论坛》《华东政法大学学报》《行政法学研究》《中国法律评论》《当代法学》《求是学刊》《社会科学》《心理学报》等中英文核心期刊发表文章四十余篇，并有部分文章被《新华文摘》《中国社会科学文摘》《高等学校文科学术文摘》《人大复印报刊资料》全文转载；在《人民日报》《光明日报（理论版）》《经济日报》《中国社会科学报》等权威媒体发表评论性文章二十余篇。

## ■ 10. 大规模图数据管理与分析论坛

### 10.1 论坛日程

地点：金鸡厅 2 会议室

时间：2025 年 8 月 2 日，星期六，下午 14:00 ~ 17:00

论坛主席：赵翔（国防科技大学）、张文涛（北京大学）

| 时间          | 报告题目                    | 报告人       | 单位             |
|-------------|-------------------------|-----------|----------------|
| 14:00-14:10 | 论坛主席致辞                  | 赵翔<br>张文涛 | 国防科技大学<br>北京大学 |
| 14:10-14:40 | 高效大图数据近似计算              | 马帅        | 北京航空航天大学       |
| 14:40-15:10 | 大规模图数据上的正则路径查询          | 王鑫        | 天津大学           |
| 15:10-15:30 | 茶歇                      |           |                |
| 15:30-16:00 | Nezha: 面向异构硬件的超大规模图处理系统 | 唐博        | 南方科技大学         |
| 16:00-16:30 | 图数据的高效压缩与查询             | 张帆        | 广州大学           |
| 16:30-17:00 | GOpt: 模块化统一图查询编译框架      | 赖龙彬       | 阿里通义实验室        |

### 10.2 论坛简介

图数据凭借其强大的表达能力，能够高效描述实体间的复杂联系，已成为工业界和学术界的研究重点，在知识图谱推理、社交关系发现、生物分子网络等众多领域有着广泛应用。随着数据规模的爆炸式增长和应用场景的不断拓展，图数据处理面临着诸多挑战：大规模图数据的计算效率有待提升，正则路径查询在实际应用中难以满足高效性需求，异构硬件环境下的超大规模图处理系统架构需要优化，图数据的压缩与查询效率需进一步提高，复杂图模式查询在多语言支持和优化技术集成方面存在瓶颈。本次论坛聚焦图数据处理领域的前沿技术与实践成果，邀请了该领域的专家学者，他们将围绕高效大图数据近似计算、大规模图数据上的正则路径查询、面向异构硬件的超大规模图处理系统、图数据的高效压缩与查询以及模块化统一图查询编译框架等关键技术展开探讨，分享最新研究进展，梳理当前存在的问题，并展望未来研究方向，以促进图数据处理领域的技术创新与发展。

## 10.3 论坛主席

赵翔



**专家介绍：**赵翔，国防科技大学大数据与决策实验室智能决策技术研究部主任，教授、博导。教育部青年长江学者，湖南省科技创新领军人才，军队青年科技英才。从事大数据知识工程、世界模型等方面的教研工作，主持国家重点研发计划青年科学家项目、国家自然科学面上基金、湖南省杰出青年基金等重要科研项目 10 余项，出版专著 2 部，获评国际学术会议论文奖 4 项，申请专利 40 余项。湖南省科技创新团队骨干成员，获评军队科技进步二等奖、中国指挥控制学会科技进步一等奖、中国计算机学会自然科学二等奖、湖南省高等教育教学成果一等奖、湖南省教学科学研究优秀成果二等奖。中国计算机学会杰出会员、数据库专委执行委员。

张文涛



**专家介绍：**张文涛，北京大学国际机器学习研究中心助理教授、研究员、博士生导师，曾任职于腾讯机器学习平台部、Apple AIML 和加拿大 Mila 人工智能实验室。研究兴趣为以数据为中心的人工智能（Data-centric AI, DCAI）、大模型和数据管理。主持国家自然科学基金青年项目、重大研究计划项目、科技部重点研发计划项目课题，近 5 年在机器学习（ICML/NeurIPS/ ICLR）、数据挖掘（SIGKDD/WWW）和数据管理（SIGMOD/VLDB/ICDE）等领域发表 CCF-A 类论文 80（一作/通讯 40+）余篇，也担任多个国际顶会的 PC Member/Area Chair。他获得多个最佳论文奖（如 WWW'22-第一作者，APWeb'23-通讯作者，CIKM'24），领导或参与开源了多个机器学习系统。他曾获 Apple Scholar、世界人工智能大会云帆奖、北京大学/北京市/中国人工智能学会优秀博士学位论文奖、未名青年学者、世界互联网大会领先科技成果奖、华为火花奖、中国电子学会科技进步一等奖等多项荣誉。

## 10.4 论坛报告

### 高效大图数据近似计算

马帅



**报告摘要：**图数据表达能力较强，成为工业界和学术界共同的重点关注之一。本报告将介绍高效大图数据计算的一种主要方法，近似计算（包括查询近似和数据近似），通过案例来介绍相关方法的思想。最后，讨论近似计算的潜在研究方向。

**专家介绍：**马帅，北京航空航天大学计算机学院教授。主要从事大数据理论与系统的研究，其成果持续发表在数据库领域、数据挖掘、人工智能和系统等顶级会议和期刊。曾任多个国际顶级期刊编委及国际顶级领域主席、程序委员。曾获国家级人才称号和 VLDB 最佳论文奖、中国电子学会科技进步特等奖和自然科学一等奖等。

## 大规模图数据上的正则路径查询

王鑫



**报告摘要：**正则路径查询（Regular Path Query, RPQ）作为图数据库中最核心的导航式查询类型之一，被广泛应用于知识图谱推理、社交关系发现、生物分子网络等场景；然而，RPQ 理论上的多项式时间复杂度在真实大规模图数据上仍难以满足交互式分析对于高效率的需求。针对这一挑战，本报告将系统回顾 RPQ 的研究脉络，梳理其理论基础；讨论大规模图数据上正则路径查询的优化方法；通过实验给出各类图数据引擎中正则路径查询的性能比较与分析，为新一代图数据管理平台上的导航式查询优化提供参考与启示。

**专家介绍：**王鑫，天津大学英才教授、博导，人工智能学院副院长。国家重点研发计划项目负责人，教育部重点领域教学资源及新型教材建设项目知识工程领域首席专家，教育部重点领域知识工程课程群虚拟教研室带头人。中国计算机学会杰出会员、中国计算机学会信息系统专委会秘书长。研究方向：知识图谱与大模型、知识数据处理。在 IEEE TKDE、SIGMOD、VLDB、ICDE、WWW、IJCAI、《计算机学报》、《软件学报》等国内外学术期刊和会议上发表论文 150 余篇。担任多个国际会议程序委员会主席；《计算机工程与应用》副主编、国际期刊 KBS、DSE、WWWJ 副主编；获天津市教学成果奖特等奖、天津市科技进步一等奖、中国计算机学会科技进步二等奖，入选全国高校计算机专业优秀教师奖励计划。

## Nezha: 面向异构硬件的超大规模图处理系统

唐博



**报告摘要:** 随着硬件技术的高速发展，越来越多的新型硬件被广泛应用于生产环境中，如 CPU、GPU 和 RDMA 等。现有的图处理系统从架构上并没有充分利用这些新型硬件的特性。在该报告中，我将结合团队近年来在图处理系统上研究积累，简要阐述架构面向异构硬件的超大规模图处理系过程中踩过的坑、取得的成绩以及收获的经验。

**专家介绍:** 唐博，南方科技大学长聘副教授/AlayaDB.AI 首席科学家，博士生导师，国自然优秀青年基金获得者。研究方向为数据库系统和大数据技术，研究成果一贯发表于数据工程与数据管理顶尖国际会议和期刊上（如 SIGMOD, PVLDB, TKDE 等），研发技术广泛应用于 Microsoft 办公软件 Excel 和开源列式存储数据库 MonetDB 系统中。他曾三次获得华为火花奖以及三次率领团队获得 CCF-A 类会议专业竞赛全球冠军（SIGMOD2021, NeurIPS2021 和 SIGMOD2024）。

## 图数据的高效压缩与查询

张帆



**报告摘要：**图数据对现实应用有强大的建模能力，能够高效地分析实体之间的复杂联系。计算机网络、社交网络等各类大规模网络，通常被模式化为图数据，并且呈现出规模宏大、关系复杂、动态变化等特点。因此，图数据的压缩和查询成为图数据研究的重要一环，能够支撑高效的数据存储、传输、计算和规模化处理。报告会重点介绍基于图总结问题的图数据压缩技术，以及面向压缩数据的查询算法研究，包括子图、路径等模块的高效计算。报告也会介绍课题组在该领域的最新进展，并且探讨未来研究方向。

**专家介绍：**张帆，广州大学网络空间安全学院教授，担任信息网络业务区负责人、大数据计算与智能研究所副所长、方滨兴院士班副点评教师、XCPD 代码训练队指导老师等职务。主要研究方向为图数据的高效管理与分析，包括子图搜索、图分解、图压缩、鲁棒性优化、传播优化等。在 SIGMOD、KDD、VLDB、SIGIR、TheWebConf、ICDE、VLDB Journal、TKDE 等国际顶级学术会议/期刊发表论文 40 余篇。作为主要负责人与阿里巴巴、华为、华南路桥等企业开展了产学研合作，主持国家及省部级项目 3 项。获得 2022 年“CCF 科技成果奖”自然科学二等奖、2023 年 ACM SIGMOD 中国新星奖等学术奖励。个人主页为 <https://fanzhangcs.github.io/>。

## GOpt：模块化统一图查询编译框架

赖龙彬



**报告摘要：**复杂图模式（CGP）将图模式匹配运算与关系运算相结合，在实际应用中有广泛应用。现有图数据库系统多采用一体化架构处理 CGP，该架构难以在单一系统中支持多种查询语言（如 Gremlin、Cypher 等），也较难集成学界提出的各类高级图查询优化技术。为此，我们提出了模块化统一图查询优化框架 GOpt，其具备三大特性：(1) 支持包括 Gremlin 和 Cypher 在内的多语言查询；(2) 解耦图查询优化器与执行引擎，具备良好的可嵌入性；(3) 集成学界提出的前沿优化技术。具体而言，GOpt 通过高层接口 GraphIrBuilder，能够将多种图查询语言统一转换为标准的中间表示（GIR），确保各种查询语言共享一套优化流程；同时提供底层接口 PhysicalConverter，负责将优化后的逻辑执行计划适配为兼容后端系统的物理执行方案，实现后端系统的无缝集成。此外，GOpt 的核心是一个原生图优化器，其整合了启发式规则集、自动类型推导方法以及针对 CGP 的代价估计技术。通过在 Neo4j 和 GraphScope 两大主流图数据库系统中的实际集成与验证，集成后，Neo4j 的平均查询速度提升 9.2 倍（最高达 48.6 倍），GraphScope 的平均查询速度提升 33.4 倍（最高达 78.7 倍）。

**专家介绍：**赖龙彬博士分别于 2010 年和 2013 年在上海交通大学获得学士学位和硕士学位。随后，他进入悉尼新南威尔士大学计算机科学与工程学院数据库组攻读博士学位，师从林学民教授和秦璐教授。他于 2017 年 3 月完成博士学业，随后加入阿里巴巴，致力于为其电商平台开发大规模图数据分析系统。他也是开源项目 GraphScope 的核心贡献者之一。他的工作成果包括：(1) GAIA：用于大规模图查询的分布式数据流系统；(2) GLogS：分布式交互式图模式匹配系统；(3) GOpt：统一图查询优化框架。目前，赖博士是阿里巴巴通义实验室的一员，在此他负责基于大语言模型（LLMs）创新应用的研究与开发工作。

## ■ 11. 时序数据智能管理与分析论坛

### 11.1 论坛日程

地点：百花厅 1 会议室

时间：2025 年 8 月 2 日，星期六，下午 14:00-17:30

论坛主席：胡吉林（华东师范大学）、宋轩（吉林大学）

| 时间          | 报告题目                                  | 报告人       | 单位             |
|-------------|---------------------------------------|-----------|----------------|
| 14:00-14:10 | 论坛主席致辞                                | 胡吉林<br>宋轩 | 华东师范大学<br>吉林大学 |
| 14:10-14:40 | 时间序列大数据计算：探索与实践                       | 王宏志       | 哈尔滨工业大学        |
| 14:40-15:10 | 面向数据库自治服务的<br>时间序列数据管理                | 陈璐        | 浙江大学           |
| 15:10-15:40 | 面向泛化与决策的下一代时空数据智能<br>建模：跨尺度、跨域与跨智能体视角 | 汪炀        | 中国科技大学         |
| 15:40-16:00 | 茶歇                                    |           |                |
| 16:00-16:30 | 超时空图神经网络                              | 徐庸辉       | 山东大学           |
| 16:30-17:00 | 时空数据紧致无监督学习与预测技术                      | 王飞        | 中科院计算所         |
| 17:00-17:30 | 基于时序智能的云平台智能巡检                        | 尹丰彬       | 阿里云计算平台        |

### 11.2 论坛简介

随着物联网和智能感知技术的快速发展，海量时间序列数据在金融、工业、气象、交通等领域持续涌现，这些数据蕴含着丰富的演化规律和商业价值。通过对时序数据的深度挖掘与分析，我们能够更准确地预测未来趋势、识别异常状态并优化决策流程。传统时序分析主要依赖于统计学习方法，如 ARIMA 和 Prophet 等基于固定假设的模型。然而，面对日益增长的时空数据规模和复杂度，这类方法已难以满足实际需求。近年来，以 MLP、RNN、GNN 和 Transformer 为代表的深度学习方法展现出强大优势，它们通过数据驱动的方式自动学习时序特征，有效捕捉长期依赖和非线性关系。特别是大模型技术的突破，为时序分析开辟了新的研究范式。本次论坛将重点探讨时序大数据计算实践、数据库自治服务中的时序管理、跨域跨尺度的时空智能建模、超时空图神经网络、紧凑无监督学习技术以及云平台智能巡检等前沿方向，旨在促进时序智能技术在理论研究与应用落地方面的深度融合，为各行业数字化转型提供核心支撑。

## 11.3 论坛主席

胡吉林



**专家介绍:** 胡吉林, 华东师范大学数据科学与工程学院, 教授, 国家级青年人才。曾担任奥尔堡大学长聘副教授, 长聘教轨助理教授, 阿联酋起源人工智能研究院担任研究员, 加州伯克利分校访问学者。研究方向主要是时空数据管理与分析, 交通分析及预测, 图神经网络。已在 CCF 推荐 A 类国际期刊或会议论文 40 余篇, 曾获得 ICDE 2022 最佳论文奖, PVLDB 2024 最佳论文奖提名, ACM SIGSPATIAL 中国新星奖。担任 IEEE Computer Society 丹麦分会主席、ACM SIGSPATIAL 中国分会执行委员、CCF 数据库专委会执行委员、CAAI 智能服务专业委员会委员和 CCF 大模型论坛执行委员。

宋轩



**专家介绍:** 宋轩教授, 国家级科技创新领军人才, 目前是吉林大学校务委员, 人工智能学院院长, “唐敖庆学者”卓越教授, 同时担任吉林省智慧警务重点实验室和吉林省超智慧城市国际科技合作重点实验室主任。主要研究方向为人工智能相关领域, 在过去 10 年间, 他在计算机领域知名的国际期刊和会议上发表各类论文 190 余篇, 其中发表在 JCR 一区或中国计算机协会推荐的 A 类期刊会议论文 140 余篇, 出版学术专著两部, 牵头制订团体标准 5 项, 申请国内国际专利 110 余项, 相关研究成果产生了巨大的社会和经济效益。

## 11.4 论坛报告

### 时间序列大数据计算：探索与实践

王宏志



**报告摘要：**时间序列在智能制造、金融、智慧城市、在线教育等诸多应用中起着至关重要的作用。对这些应用中时间序列进行有效计算对这些领域的成功起着至关重要的作用。时间序列大数据计算带来了数据强关联、模式复杂、有序时间窗口等诸多挑战。这个报告将从大数据计算角度介绍时间序列大数据管理、治理和分析等计算带来的挑战和报告人团队在这个领域从理论到应用研究成果，包括时序数据管理、多种时间序列分析算法和时间序列清洗算法与系统等。

**专家介绍：**王宏志，哈尔滨工业大学计算学部长聘教授、博士生导师、中国计算机学会杰出会员、数据库专业委员会常务委员、大数据专家委员会委员、数据治理发展委员会执行委员、黑龙江省大数据科学与工程重点实验室主任、哈工大海量数据计算研究中心主任、黑龙江省计算机学会数据科学与大数据技术专委会主任，哈工大数据库与大数据技术专业负责人。主要从事数据库和大数据研究，主持国家自然科学基金重点项目和联合基金项目各 1 项，各类国家、省部级和校企合作项目 20 余项，在 VLDB, SIGMOD 等国内外重要会议和期刊发表学术论文 300 余篇，出版学术专著四部，论文被 SCI 收录百余次，他引 6000 余次，授权发明专利 50 余项，获得省部级一等奖 2 项。其研究成果应用于国产数据库、电力系统、智慧金融等领域。个人主页:[homepage.hit.edu.cn/wang](http://homepage.hit.edu.cn/wang)。

## 面向数据库自治服务的时间序列数据管理

陈璐



**报告摘要：**时间序列数据对实现高效的数据库运维与自主管理具有关键作用。通过对数据库相关时间序列数据（如性能指标、SQL 日志等）实施存储优化、异常检测、预测分析等时序管理技术，可显著提升系统自动化水平并降低人力成本。本报告浅谈讲者对于面向数据库自治服务的时间序列数据管理技术发展的思考，而后介绍讲者负责团队近年来在这方面的进展。

**专家介绍：**陈璐，浙江大学百人计划研究员，博士生导师，国家特聘青年专家（2021）。研究方向为数据库和大数据处理。已在国内外顶级学术期刊或会议 VLDBJ、TKDE、SIGMOD、VLDB、ICDE、SIGIR 等发表论文 60 余篇，出版中文学术专著 1 部，授权发明专利 10 项，登记软件著作权 3 项，并获 DASFAA 2024 最佳论文奖、ICDE 2022 最佳论文奖、ICDE 2019 优秀论文（One of the Best Papers）、APWeb-WAIM 2018 最佳论文奖、2017 年度 CCF 优秀博士学位论文奖、ICDE 2015 优秀论文等。

## 面向泛化与决策的下一代时空数据智能建模： 跨尺度、跨域与跨智能体视角

汪炀



**报告摘要：**报告围绕多源多尺度时空数据的智能建模与泛化预测，系统阐述了课题组在时空数据基础建模、分布外泛化、小大模型协同以及赋能重大科研仪器装备等方面的最新研究进展。针对时空建模基础理论，我们提出跨变量跨时间尺度的纯卷积建模与常微分可微建模框架，解决多元异构时空建模、多频率时空数据高效预测及模型边缘部署难题；针对模型泛化挑战，提出基于神经结构与认知启发构建进化式学习体系，增强模型的跨域迁移与持续泛化能力；同时探索大模型与小模型的协同机制，提升智能体系统在时空智能与复杂决策中的智能水平。最后，我们将结合“智能科学家”与“红外芯片科学家”等典型案例，展示 AI for Science 范式下时空（时序）智能对重大科研仪器装备的赋能潜力，并展望 NeuroAI 与科学智能驱动下的下一代时空建模图景。

**专家介绍：**汪炀，中国科学技术大学教授、博导、软件学院副院长，CCF 物联网专委会委员、智慧交通分委会执行委员。长期致力于时空智能基础理论及面向物质科学的人工智能交叉前沿研究。以第一/通讯作者身份发表中科院一区及国际顶级学术会议在内的高水平学术论文 100 余篇，其中代表性成果发表在 Nature 子刊、PNAS、JACS、IEEE TPAMI 等顶级期刊及 ICML、NeurIPS、ICLR、KDD、Mobicom 等顶级国际学术会议，论文曾入选 FESE 创刊 10 年 10 篇代表性论文，并六次入选顶级会议 Oral/Spotlight Paper。作为项目/课题负责人承担中国科学院科技制高点专项、国家重大科研仪器研制项目（部门推荐）、国家发改委下一代互联网专项、国家自然科学基金项目等 30 余项国家级科研项目，累计项目经费超过 1.5 亿元，曾获 IBM 全球杰出学者奖、安徽省杰出青年基金、中国科学院年度团队、安徽省优博/ACM SIGSPATIAL 优博导师奖等奖励。

## 超时空图神经网络

徐庸辉



**报告摘要：**现实世界中的众多复杂系统，从城市交通网络、社交信息传播、到生物分子相互作用和物理系统演化，本质上都同时具备图结构（实体及其关系）和时空动态性（状态随时间、空间演变）。传统的图神经网络在捕捉静态图结构方面表现出色，但在处理复杂的时空依赖关系、尤其是跨时空尺度的交互时，仍面临巨大挑战。本次报告聚焦于超时空图神经网络前沿方向，通过新颖的架构设计，融合注意力、时空卷积、图扩散等，实现时空特征的深度融合与联合学习，直接刻画图结构本身的动力学演化，并利用超图等机制有效表征高阶时空交互。报告将阐述其关键设计思想、典型架构，并通过实际应用展示其相比传统模型的显著优势。

**专家介绍：**徐庸辉，山东大学教授，博士生导师，主要研究可信人工智能、多模态大模型、图神经网络等。在国际人工智能顶级会议 AAAI、IJCAI、ICDE 及权威期刊 TNNLS、TKDE、TKDD 等发表论文 100 余篇，申请授权核心知识产权 20 余项。长期担任人工智能与数据挖掘领域顶级学术会议 IJCAI、AAAI、KDD 等程序委员，权威学术期刊 IEEE TKDE、TNNLS 等审稿人。先后承担国家重点研发计划课题、国家自然科学基金项目、国家自然科学基金重大项目子课题、阿里巴巴全球创新研究计划等国家级省部级项目等 20 余项，参与获得国家级教学成果奖二等奖、山东省科学技术进步奖一等奖、山东省教学成果奖特等奖。

## 时空数据紧致无监督学习与预测技术

王飞



**报告摘要：**时空数据深度学习技术的发展，给诸多应用领域的智能化预测、决策与优化提供了丰富的想象力。然而由于时空数据生成机制互异，难以约简深度学习模型的一般寻优过程，导致基于深度学习提取有效信息并进行预测已经成为时空数据分析的共性难题之一。本报告首先阐述时空数据在走向通用人工智能中的重要意义，然后聚焦时空数据表征的尺度适应问题、本征时空依赖的挖掘优化问题、跨域时空预测的异质性建模问题，详细介绍在预训练增强框架、时空身份编码简约架构、跨域模型自适应重建等方面的相关进展，旨在为时空数据深度学习模型的多领域落地提供发展启示。

**专家介绍：**王飞，博士，副研究员。专注时空数据挖掘、时序数据预测、时空 AI for Science 等方向研究，在 KDD、VLDB、TKDE、The Innovation 等领域重要学术会议和期刊上发表学术论文 50 多篇，曾获 DASFAA 2022 最佳学生论文奖、The Innovation 期刊最佳论文奖 2 次、城市科学大会最受欢迎墙报奖，STEP、GinAR、STID、DSFormer 等多篇论文入选最有影响力论文榜单，谷歌引用超 4200 次。担任 The Innovation 青年编委，IEEE CIS 时序时空数据方向 Task Force 首批程序委员会成员。先后入选中国科学院计算所“新百星”、中国科学院“青年创新促进会”、中国科学院特聘研究岗位（骨干）、中国科学院“西部之光”交叉团队。指导/协助指导学生获国家奖学金 3 次，计算所所长特别奖（夏培肃奖）1 次，中国科学院优博 1 次。承担基金委、科技部等项目 20 余项，在多个用户单位取得重要应用成果，曾受科技部等重要用户多次书面感谢。获中国指控学会科技进步奖一等奖、二等奖。

## 基于时序智能的云平台智能巡检

尹丰彬



**报告摘要：**云平台智能巡检的核心挑战在于如何从高维、多源、动态变化的时序数据中提取有效信息。本次分享将探讨我们为此构建的一套时序智能引擎。该引擎通过统一的异常检测模型克服“告警孤岛”，并引入聚集度、新奇度、风险程度等多个机制量化系统状态；同时，结合历史案例和大模型推理能力，实现从风险预警到问题定界的自动化闭环。我们将结合真实案例，剖析算法选型、模型优化及工程落地中的关键问题与解决思路，为同行提供可借鉴的 AIOps 实践经验。

**专家介绍：**尹丰彬，阿里云高级算法工程师，毕业于西安交通大学少年班致力于 AIOps 领域的研究与实践，曾主导设计并研发了面向大数据产品的智能问答系统，降低了产品答疑人力成本。在时序数据分析领域发表相关专利 2 项，曾在 CIKM 上分享根因定位相关成果，目前专注于将大模型和前沿的时序智能算法相结合，并在实际生产场景中落地，以数据驱动的方式解决云平台复杂因素带来的不稳定性。

## ■ 12. 数据库系统新技术论坛

### 12.1 论坛日程

地点：百花厅 2 会议室

时间：2025 年 8 月 2 日，星期六，下午 14:00 ~ 17:30

论坛主席：柴云鹏（中国人民大学）、卢卫（中国人民大学）

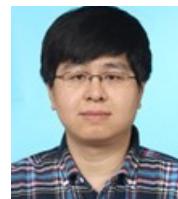
| 时间          | 报告题目                    | 报告人                           | 单位  |
|-------------|-------------------------|-------------------------------|---|
| 14:00-14:10 | 论坛主席致辞                  | 柴云鹏<br>卢卫                     | 中国人民大学                                    |
| 14:10-14:40 | 神经符号数据库技术及应用            | 袁野                            | 北京理工大学                                    |
| 14:40-15:10 | 大模型时代的数据管理技术：<br>进展与展望  | 高云君                           | 浙江大学                                      |
| 15:10-15:30 | 茶歇                      |                               |   |
| 15:30-16:00 | 三层解耦云原生数据库<br>关键挑战与技术创新 | 李飞飞                           | 阿里云                                       |
| 16:00-16:30 | 数据库研究中的<br>一些冷门却仍然重要的问题 | 周烜                            | 华东师范大学                                    |
| 16:30-17:00 | 大模型赋能的数据库自优化技术          | 张心怡                           | 中国人民大学                                    |
| 17:00-17:30 | Panel:                  | 袁野<br>高云君<br>李飞飞<br>周烜<br>张心怡 | 北京理工大学<br>浙江大学<br>阿里云<br>华东师范大学<br>中国人民大学 |

### 12.2 论坛简介

近年来，随着新型硬件的发展、人工智能的融入以及系统架构的持续演进，数据库系统技术呈现出快速演化的趋势。数据库系统前沿技术论坛聚焦当前数据库系统领域的关键研究方向与创新成果，涵盖向量数据库、云数据库、AI4DB（人工智能赋能的数据库系统）以及神经符号数据库等前沿主题。向量数据库支撑大模型检索与多模态应用，云数据库推动系统向弹性、可管理方向演进，AI4DB 利用 AI 提升数据库自动化能力，神经符号数据库探索数据理解与推理的深度融合。论坛致力于为学术界与产业界构建高水平的交流平台，促进理论研究与系统实践的协同发展，推动数据库系统向智能化、云端化、融合化方向持续演进。

## 12.3 论坛主席

柴云鹏



**专家介绍：**柴云鹏，中国人民大学教授、博士生导师、中国人民大学“吴玉章学者”特聘教授，目前担任信息学院副院长（主持工作）、党委副书记。中国计算机学会教育工委委员、数据库专委会执行委员、信息存储专委会执行委员，主要研究方向为数据库系统、存储系统、云计算、智能计算系统。研究成果发表在 SIGMOD、ASPLOS、DAC、ICDE、TOS、TPDS 等国际顶级会议期刊。获得深圳市科技进步一等奖、北京市教学成果一等奖、中国产学研创新与促进奖产学研合作创新成果奖一等奖、中国计算机学会科技进步二等奖等奖励。

卢卫



**专家介绍：**卢卫，中国人民大学教授，信息学院副院长，博导，中国计算机学会数据库专委执行委员，全国大学生计算机系统能力大赛数据库赛道技术负责人。近年来主要从事数据库基础理论、分布式数据库系统等相关领域研究，在 SIGMOD、VLDB 等 CCF A 类会议或期刊发表论文 40 余篇，曾主持国家重点研发计划课题、国家自然科学基金项目、北京市重大专项课题等。其研究注重成果落地，长期与数据库领域的头部企业保持紧密合作，多项研究成果已落地到 TDSQL、Kingbase 等数据库产品中。主讲的课程《数据库系统概论》获得国家级线上线下混合式一流本科课程和国家级线上一流本科课程。曾获北京市高等教育教学成果一等奖、深圳市科技进步一等奖、首届华为云难题“火花奖”等。

## 12.4 论坛报告

### 神经符号数据库技术及应用

袁野



**报告摘要：**神经符号数据库（Neuro-Symbolic Database），作为一种新型的数据库平台，深度融合了人工智能分析、数据库查询以及大数据计算的关键技术。其突出优势在于不但能够高效地处理异构数据，包括结构化和半结构化数据，还能够有效地管理和解析非结构化数据，如图像、视频和文本等，实现对多源异构数据的深度挖掘。本报告将深入探讨神经符号数据库在处理各类复杂数据类型和模式下的应用潜力，旨在揭示其在数据密集型领域的广阔应用前景，并对神经符号数据库面临的技术挑战和我们的解决思路进行介绍，包括神经符号数据库的查询语言、处理算法和存储方案等。总体来看，本报告以剖析神经符号数据库的关键技术问题为核心，力求为该领域的进一步研究和发展提供有力的理论支撑和实践建议。

**专家介绍：**袁野，北京理工大学教授、博士生导师、国家杰青和优青基金获得者。主持国家自然基金重点项目，科技部重点研发项目。曾获国家科技进步二等奖，中国电子学会自然一等奖，教育部和辽宁省科技进步一等奖、全国优秀博士论文提名奖、中国计算机学会优秀博士论文奖。袁博士是中国人工智能学会服务专委副主任、中国计算机学会数据库专业委员会副主任、IEEE、ACM 高级会员。主要研究方向为人工智能、大数据。发表相关CCF A类论文200余篇。

## 大模型时代的数据管理技术：进展与展望

高云君



**报告摘要：**在大模型时代，人机交互正经历从“机械指令”到“自然对话与多模态融合”的变革，对底层的数据管理技术提出了新要求。“自然对话与多模态融合”的人机交互模式涉及大量的非结构化数据，并需要进行多模态数据的查询与分析，因而亟需破解大规模非结构化数据存储，高效多模态数据查询，自动化数据分析等难题。鉴于此，报告人负责团队已在向量数据库、新硬件架构、多源异构数据查询、Text-to-SQL、检索增强生成、数据分析智能体等方面做了一系列探索。本报告将先汇报其相关进展，而后展望可能的发展方向。

**专家介绍：**高云君，浙江大学求是特聘教授，博士生导师，国家杰出青年科学基金获得者（2020）、国家优秀青年科学基金获得者（2015）。研究方向为数据库、大数据管理与分析、DB 与 AI 融合，已发表 CCF A 类论文 180 余篇，出版专著或教材 6 本，授权专利 40 余项，登记软著 10 余项，并获 CCF A 类论文 VLDB、ICDE 最佳或优秀论文等 7 次，省部级或全国性学会科技进步特等或一等奖 3 项。现为 ACM SIGSPATIAL 中国分会副主席，全省大数据智能计算重点实验室主任，浙江大学软件学院常务副院长。担任 DSE 期刊执行主编，TKDE、JCST、FCS、《计算机研究与发展》等期刊（青年）编委/副编辑（Associate Editor），VLDB、ICDE、DASFAA 等 10 余个顶级或重要会议程序委员会/Workshop/Tutorial/Demo/宣传/出版/本地联合主席（Co-Chair）或领域主席（Area Chair），SIGMOD、VLDB、ICDE、SIGKDD、SIGIR 等会议（资深）程序委员会委员。

## 三层解耦云原生数据库关键挑战与技术创新

李飞飞



**报告摘要：**云原生数据库向纵深发展需要在数据中心规模实现冯诺依曼架构，核心挑战是将计算、内存、存储进行独立调度和管理，实现规模性的三层解耦和池化，在保障事务一致性、数据读写一致性的关键要求下突破上述关键挑战，并同时提供高性能、高并发、高弹性、高可用的系统能力。为实现这一目标，需要系统架构设计、软硬协同、关键算法和协议、以及工程实现多维度协同和突破。本报告将系统性的介绍相关工作，在 SIGMOD 2024 和 SIGMOD 2025 连续两年获得工业界唯一最佳论文奖。最后，智能化是确保在这个尺度规模下高效运维、根因定位、以及大规模数据价值挖掘的必要手段，报告也将总结云原生数据库与智能化相结合的相关工作，实现 data+ai 的创新与落地。

**专家介绍：**李飞飞，ACM/CCF/IEEE Fellow(会士)，阿里云副总裁/高级研究员。获数据库与大数据系统方向多项国际顶级学术科研与技术会议最佳论文/成果奖(ACM SIGMOD 2025 Industry Track Best Paper Award, IEEE ICDE 2024 Industry and Application Best Paper Award, ACM SIGMOD 2024 Industry Track Best Paper Award, ACM SIGMOD 2023 Beset Paper Award, EDBT 2022 10 Years Test of Time Award, IEEE ICDCS 2020 Best Paper Award、ACM SoCC Best Paper Award Runner Up、ACM SIGMOD 2016 Best Paper Award, ACM SIGMOD 2015 Best System Demonstration Award, IEEE ICDE 2014 10 Years Most Influential Paper Award, IEEE ICDE 2004 Best Paper Award), 作为第一完成人获得世界互联网大会 2019 全球领先科技成果奖、浙江省科技进步一等奖、中国电子学会科技进步一等奖、中国计算机学会科技进步一等奖等。担任多个国际及国内一流学术期刊和学术会议的编委、主席，中国计算机学会 CCF 大数据专家委员会、数据库专业委员会常委。带领团队研发了以云原生数据库 PolarDB 为核心的阿里云瑶池数据库产品矩阵，实现了中国数据库市场份额第一，作为国内唯一数据库厂商连续 5 次进入 Gartner 全球 Cloud DBMS 市场分析报告领导者象限，登顶 TPC-C 官方性能和性价比全球第一。

## 数据库研究中的一些冷门却仍然重要的问题

周烜



**报告摘要：**我们很重视新的应用、技术或硬件为数据库领域带来的发展机遇，但较少讨论数据库系统在软件体系中的实际地位正在发生什么变化。这次聊一聊这个方面的课题。

**专家介绍：**周烜，华东师范大学特聘教授、数据科学与工程学院副院长。2005 年于新加坡国立大学取得博士学位，先后在德国 L3S 研究中心、澳大利亚联邦科工组织（CSIRO）和中国人民大学信息学院任职，2017 年入职华东师范大学。研究工作聚焦数据库系统领域，曾负责多个国内外的科研项目，与工业界头部科技企业保持长期深入合作，多项技术成果在工业界落地应用，曾获国家科技进步二等奖、国家级教育成果二等奖，入选国家高层次人才计划。

## 大模型赋能的数据库自优化技术

张心怡



**报告摘要：**随着数据库系统复杂性和数据规模的持续增长，传统依赖规则和固定模型的优化方法正逐渐暴露出局限性。近年来，以生成式语言模型为代表的大模型技术飞速发展，使机器理解各行各业的数据、业务流程和复杂用户意图成为可能。在大模型的加持下，Agent 技术进一步推动了复杂交互环境下的认知与自主决策能力，为数据库自优化提供了新的技术支撑和研究视角。另一方面，随着国产化数据库和开源数据库的蓬勃发展，数据库系统源代码的开放性正在为突破传统黑盒调优范式提供重要契机。通过大模型对数据库内部机制、查询执行计划、系统日志等多模态信息的理解与推理，有望实现从黑盒到“灰盒+智能”结合的自优化路径。本报告将梳理了大模型赋能数据库自优化的最新进展与面临的关键挑战，并汇报大模型在数据库系统理解、优化策略生成和自适应调优等方面的研究进展与探索。

**专家介绍：**张心怡，中国人民大学助理教授，吴玉章青年英才，CCF 数据库专委委员，长期从事数据库系统、AI4DB（人工智能赋能的数据库系）方向的研究，在该领域发表 CCF A 类论文十余篇，入选 2025 年“高创计划”青年人才托举工程、2025 年中国计算机学会数据库专委会博士学位论文激励计划、2023 年字节跳动奖（全球仅 13 人）、2022 年微软学者提名（亚太地区仅 21 人）、北京大学优秀博士论文、NDBC 最佳学生论文奖、。主持 CCF-阿里云瑶池基金、CCF-华为胡杨林基金，担任 VLDB、TKDE、TODS、JCST 等会议及期刊的审稿人，并担任 ICDE、WWW 程序委员会成员。

## 12.5 Panel

### 数据库系统前沿技术与智能化演进

柴云鹏、卢卫、袁野、高云君、李飞飞、周烜、张心怡

**简介：**聚焦数据库技术在人工智能深度融合、云原生架构创新等多重驱动下的演进趋势，本 Panel 将围绕以下热点议题展开讨论：大模型与数据库系统结合对数据管理、优化和智能化运维带来的机遇与挑战？云原生与新硬件架构下数据库系统的关键技术突破与产业落地路径？面向未来，数据库研究中被忽视但仍重要的核心问题？

## ■ 13. CCF-华为胡杨林基金数据库专项论坛

### 13.1 论坛日程

地点：金熊厅会议室

时间：2025年8月2日，星期六，下午14:00~18:20

论坛主席：张岩峰（东北大学）、李士福（华为技术有限公司）

| 时间          | 报告题目                    | 报告人 | 单位         |
|-------------|-------------------------|-----|------------|
| 14:00-14:10 | 论坛主席致辞                  | 张岩峰 | 东北大学       |
| 14:10-14:30 | 基于对象存储的存算分离数据库数据结构      | 许建秋 | 南京航空航天大学   |
| 14:30-14:50 | 通用解耦OLTP优化器的高效描述接口      | 张心怡 | 中国人民大学     |
| 14:50-15:10 | 查询资源预估和智能任务调度           | 周煊赫 | 上海交通大学     |
| 15:10-15:30 | 面向场景化机器学习任务的高效大数据存储格式设计 | 王宏志 | 哈尔滨工业大学    |
| 15:30-15:50 | 面向NPU异构计算资源的软硬协同调度算法    | 王胜  | 武汉大学       |
| 15:50-16:10 | 结构与非结构化数据关联知识发掘         | 于明鹤 | 东北大学       |
| 16:10-16:30 | 结构与非结构化数据关联知识发掘         | 沈玮  | 南开大学       |
| 16:30-16:50 | 基于大语言模型和轻量模型结合的多模态数据分析  | 汤南  | 香港科技大学（广州） |
| 16:50-17:10 | 基于大语言模型和轻量模型结合的多模态数据分析  | 商砾  | 电子科技大学     |
| 17:10-17:30 | 统一多模态数据融合检索技术           | 张美慧 | 北京理工大学     |
| 17:30-17:50 | 统一多模态数据融合检索技术           | 王鑫  | 天津大学       |
| 17:50-18:10 | 统一多模态数据融合检索技术           | 郑卫国 | 复旦大学       |
| 18:10-18:20 | 论坛主席致谢                  | 李士福 | 华为技术有限公司   |

## 13.2 论坛简介

CCF-华为胡杨林基金数据库专项自 2019 年创立，得到国内外学者广泛关注和踊跃申报。截止当前，数据库专项总计投入逾 2000 万，支持近 50 位学术界优秀学者投身数据库前沿课题的研究，解决产业界面临的挑战难题，探索数据库系统核心技术的发展方向。2024 年签署下一个五年战略合作框架协议，在过往五年的基础上进一步加大投入，支持更多优秀学者开展前沿探索和研究。2025 年度围绕智能数据管理、云原生数据库、软硬结合系统等方向，计划开展 12 项合作项目，本论坛针对 12 项合作项目进行项目工作计划报告，来自 12 所高校的研究人员和企业界专家进行讨论。

### 13.3 论坛主席

张岩峰



**专家介绍:** 张岩峰，东北大学教授，博士生导师，东北大学与美国麻省州立大学联合培养博士，国家级高层次青年人才。研究方向为数据库系统、机器学习系统。承担国家重点研发计划课题、国家自然科学基金等国家级科研项目，承担华为 2012 实验室、阿里巴巴达摩院等校企合作项目。曾获中国计算机学会自然科学二等奖、云计算国际会议 ACM SOCC 杰出论文奖等奖励。

李士福



**专家介绍:** 李士福，GaussDB 数据库技术专家，15 年数据库内核开发经验，2014 年加入华为高斯部，曾在 Teradata 研发中心从事分布式 MPPDB 数据库研发。目前其主要负责 GaussDB（含 openGauss）自治数据库内核的研发工作，开发产品包括 DBMind 自治运维系统、DB4AI 库内 AI 引擎、ABO 优化器、向量数据库等已广泛应用于金融、政企等关基行业；同时作为 openGauss 社区 TC 成员支撑 openGauss 社区拓展。

## 13.4 论坛报告

### 基于对象存储的存算分离数据库数据结构

许建秋



**报告摘要：**在云存储服务多样化的背景下，云数据库系统需要动态弹性地获取具有差异化成本和访问特性的存储资源，并基于数据访问模式实现智能化的分层存储管理。其主要面临的挑战包括对象存储分块粒度选择、元数据存储与管理和读写局限性。本项目拟设计和实现基于对象存储的存算分离数据库数据结构，其作为中间层解决段页式存储与对象存储的异构性问题。在数据库系统中，冷数据的随机点查询、范围查询、随机修改的性能不低于热存储的 50%；范围查询、大批量修改的性能不低于热存储的 90%。

**专家介绍：**南京航空航天大学教授，计算机系主任，ACID 研究所所长，主要从事数据管理及数据库原型系统研究，近年来主要针对大模型驱动的复杂数据管理和高效计算、垂直领域智能数据管理和分析开展工作，主持国家自然科学基金项目 3 项、省部级项目 4 项。发表学术论文 50 余篇，包括 CCF 推荐 A 类论文和中科院 1 区论文等，例如 IEEE TKDE、ICDE、PVLDB、计算机学报等，出版英文专著 1 本。担任中国科学卓越期刊二期领军、中科院 2 区期刊 Frontier of Computer Science 青年编委，多次获得国际/国内学术会议论文奖，包括最佳系统演示论文奖，最佳展望论文奖等，国内学术会议特邀报告 5 次，授权国家发明专利 7 项。

## 通用解耦OLTP优化器的高效描述接口

张心怡



**报告摘要:** GaussDB 支持多种数据库生态，每个生态都有独立 SQL 引擎，为实现全生态性能优势，需构建统一的 SQL 语义和执行计划抽象描述语言，支持优化器基于通用的抽象语义来进行逻辑优化和物理优化，驱动各生态的执行器调用 GaussDB 融合引擎。现有开源查询描述语言（如 ORCA、Substrait）在抽象表达上已有探索，但在完备性、扩展性和转换效率方面不足，难以满足高频短查询的 OLTP 需求。因此，本报告关注通用解耦 OLTP 优化器的高效描述接口，探究二进制零拷贝数据转换机制，实现轻量级转化。

**专家介绍:** 中国人民大学讲师，吴玉章青年英才，CCF 数据库专委执行委员。博士毕业于北京大学，主要研究方向为智能数据库系统、面向人工智能的数据管理，在 SIGMOD、VLDB 等数据库领域国际会议和期刊上发表 CCF A 类论文 10 余篇，获 2024 年玉兰青年学者、2023 年字节跳动奖学金（全球 13 人）、2022 年微软学者提名（亚太地区 21 人）等荣誉。担任 ICDE、WWW 等国际会议 PC Member，任 VLDB、TKDE、TODS 等国际会议期刊审稿人。

## 查询资源预估和智能任务调度

周煊赫



**报告摘要：**本项目研究面向HTAP数据库的混合负载资源估算与任务调度。首先，我们将介绍面向Stage-based的细粒度、低开销的资源预测方法，准确估算查询执行的关键资源消耗。其次，我们将介绍基于pipeline执行模式的智能调度策略，实现Stage级别的资源优化与任务排序。最后，我们将介绍原型系统设计以及混合负载基准测试方案，系统评估调度方案在吞吐性能、资源隔离与可扩展性方面的表现。

**专家介绍：**上海交通大学计算机学院院长聘轨助理教授，博士生导师。主要研究智能数据分析、AI数据底座、自治数据库系统。在 SIGMOD、VLDB、NIPS、TKDE 等 CCF A 类会议和期刊上已发表论文数十篇，包括近五年 VLDB、ICDE 高被引论文，入选 CMU、Cornell 等高校课程。谷歌学术引用量两千余次。曾获通信学会科学技术一等奖、ACM SIGMOD Jim Gray 提名奖（国内首次）、VLDB 最佳工业论文亚军奖（第一作者）、CCF 优博、微软学者、清华特奖等荣誉。曾主导或主要参与 OpenMLDB（入选 Gartner 报告）、DB-GPT（BenchCouncil 开源成就奖）、Bird-SQL（用于 OpenAI 微调服务推广）等多个高星标开源项目。

## 面向场景化机器学习任务的高效大数据存储格式设计

王宏志



**报告摘要：**随着智能制造的发展，企业积累了海量结构化数据。在质量预测任务中，机器学习模型（如深度神经网络）依赖这些历史数据进行训练与推理。然而，高维度、大体量、复杂访问模式和动态变化等数据特性，使得传统存储系统在吞吐量和访问延迟上难以满足需求，严重制约了模型性能与应用部署效率。为了解决如上挑战，本项目将从两方面展开研究：1) 面向场景化机器学习任务的高维向量存储；2) 面向场景化机器学习任务的动态分层压缩，通过设计页内组织、页间布局及压缩编码等存储格式优化方法，构建面向场景化机器学习任务的高效大数据存储原型系统。

**专家介绍：**哈尔滨工业大学计算学部长聘教授、博士生导师、海量数据计算研究中心主任、哈工大数据科学与大数据技术专业负责人、CCF 杰出会员、CCF 数据库专委常务委员、大数据专家委员会执行委员、数据治理发展委员会执行委员、黑龙江省大数据科学与工程重点实验室主任、黑龙江省计算机学会数据科学与大数据技术专委会主任、黑龙江省教学名师。主要从事数据库和大数据研究，主持国家自然科学基金重点项目和联合基金项目各 1 项，各类国家、省部级和校企合作项目 20 余项，在 VLDB, SIGMOD 等国内外重要会议和期刊发表学术论文 300 余篇，出版学术专著四部，论文被 SCI 收录百余次，他引 6000 余次，授权发明专利 50 余项，获得省部级一等奖 2 项。其研究成果应用于国产数据库、电力系统、智慧金融等领域。

## 面向NPU异构计算资源的软硬协同调度算法

王胜



**报告摘要：**大语言模型（LLM）的规模化应用对推理效率提出了更高要求，尤其在资源受限场景下，如何平衡计算性能和响应速度成为关键挑战。我们聚焦 NPU（神经网络处理器）的架构优势，探索其在加速大模型推理中的潜力，并提出细粒度的资源调度策略。区别于传统依赖通用计算单元的方法，我们通过挖掘 NPU 的并行计算特性，实现计算资源的动态优化配置。同时，针对检索增强生成（RAG）技术中高计算开销的检索环节，我们分析其资源需求特征，构建分层调度机制以提升整体效率。我们的工作主要涵盖：1) NPU 的硬件加速原理及其与大模型推理的适配性；2) RAG 任务的计算特征分析与调度优化方法；3) 面向实际业务场景的系统级优化方案与性能评估。通过资源感知的调度优化，我们为高负载场景下的大模型高效推理提供系统化解决方案。

**专家介绍：**武汉大学计算机学院教授，挂任武大信息中心副主任和聘任武大数智教育教学研究中心成员；担任 CCF 武汉会员中心秘书长、数据库专委会、信息系统专委会执行委员；2013 年和 2016 年本硕毕业于南京航空航天大学，2019 年博士毕业于澳大利亚皇家墨尔本理工大学，2021 年在美国纽约大学完成博士后；主要研究一体化数据管理及应用，研究成果发表在 SIGMOD 和 PVLDB 等 CCF A 类推荐的国际顶级会议及期刊上；培养学生获国家奖学金、优秀毕业设计；担任武大数智教育课程《数据分析与处理（Python）》课程组组长、雷军班和弘毅班学业导师；主持国家重点研发计划课题、湖北省重点研发计划项目和国家自然科学基金青年基金等，入选国家和省市等各高层次青年人才项目。

## 结构与非结构化数据关联知识发掘

于明鹤



**报告摘要:** 随着大数据时代的到来，企业资源规划、工业制造等领域的数据资源从单一类型演变为结构化与非机构化共存的复杂模态。为破解两类数据割裂导致的语义失联与关系湮没等瓶颈，本项目拟研究基于知识图谱的结构化与非机构化数据关联知识发掘技术。项目首先构建知识图谱对异构数据进行统一管理，并在此基础上，通过研究跨模态知识图谱补全、多模态数据融合和可解释分类三个方向来解决结构化与非结构化数据关联知识发掘问题。从而有效梳理数据关系，融合多源数据，并精准的提取知识，破解异构数据协同分析的难题，为企业管理、工业制造、金融等领域提供实践工具。

**专家介绍:** 东北大学软件学院副教授、硕士生导师，软件工程系副主任。CCF 高级会员、CCF 数据库专委会、信息系统专委会委员。研究方向为数据库、大数据分析、智慧教育等。承担或参与国家自然科学基金、国家重点研发计划等多项国家级科研项目，及华为、OceanBase 等校企合作项目。在 VLDB、ICDE、TKDE、VLDBJ 等国际重要会议和期刊上发表论文 40 余篇。

## 结构与非结构化数据关联知识发掘

沈玮



**报告摘要：**在大数据时代，结构化数据和非结构化数据广泛存在于各个领域，它们的数据量呈现出爆发式增长。结构化数据如知识图谱、数据库表格等，结构清晰规整；非结构化数据像文本、图像等，内容丰富但格式松散。二者独立分析时，挖掘的信息有限。如能将这两类数据关联起来，打破数据类型的壁垒，定能从两类数据的关联中发掘出新的知识和规律，为企业决策、科研创新等提供更全面深入的信息支持。本项目面向企业资源管理应用场景，针对多源数据融合难、知识提取准确性低等挑战，研究结构与非结构化数据关联知识发掘任务。

**专家介绍：**南开大学计算机学院教授、博士生导师，CCF 数据库专委会执行委员。博士毕业于清华大学计算机系，主要研究方向是数据融合，以第一/通讯作者在数据库、数据挖掘领域 CCF-A 类国际会议期刊（SIGMOD、KDD、ICDE、WWW、TKDE 等）上发表论文 20 余篇，所发表一作/通讯论文被引用超过 2000 次。先后主持国家自然科学基金面上项目、重点项目课题、青年项目，以及 CCF-滴滴盖亚学者科研基金、CAAI-华为 MindSpore 学术奖励基金等多项产学研项目。曾获中国科协青年人才托举工程、ACM 中国新星奖提名奖、CCF-Intel 青年学者提升计划、中国人工智能学会优秀博士论文奖、天津市青年科技优秀人才。

## 基于大语言模型和轻量模型结合的多模态数据分析

汤南



**报告摘要：**本研究聚焦多模态数据分析系统的智能协同计算范式，致力于解决非结构化数据爆炸与模态多样性（文本、图像、音视频等）带来的核心挑战。传统方法受限于规则系统的僵化与大模型的高昂计算成本，尤其在跨模态语义融合、实时性要求及能效控制方面存在显著瓶颈。本领域通过重构“解析-调度-执行-反馈”的技术链条，构建分层智能框架，旨在突破“算力堆砌”模式，实现多模态数据从“暗数据”到商业洞见的高效转化。本研究领域的核心创新在于构建大语言模型（LLM）与轻量模型的动态协同机制：通过**任务解耦技术**将复杂多模态分析分解为全局语义规划（由LLM执行高阶推理）与局部信息抽取（由轻量语言模型完成），显著降低计算密度；依托**马尔可夫决策过程（MDP）的动态调度框架**实现异构资源优化分配，在保障语义完整性的前提下提升系统吞吐量；创新性采用**分组投影对齐策略**解决跨模态特征空间碎片化问题，在专业领域使细粒度推理提升准确率，最终形成支撑TB级数据实时处理的高效引擎，为工业场景提供兼具强认知能力与低资源消耗的智能化分析基础设施。

**专家介绍：**香港科技大学（广州）副教授。2007年12月取得香港中文大学系统工程与工程管理专业博士学位。2008-2009年期间，任职研究院于荷兰国家数学与计算机研究所，主要研究方向是列存储数据库对半结构化数据的高效支持。2010-2011年期间，任职研究院于爱丁堡大学，主要研究方向为数据清洗。2011年12月至2023年6月，加入卡塔尔计算机研究所担任资深科学家，作为卡塔尔国家计算机研究所的创始成员，在过去十年的时间里带领卡塔尔国家计算机研究所成为数据管理领域世界闻名的研究中心。在数据库和数据挖掘顶级会议和期刊上发表论文150余篇。根据Google Scholar，其论文总引用数量超过9000，h-index为49。在众多国际顶级会议和期刊担任审稿人，区域主席(VLDB 2026 & 2025, ICDE 2026 & 2024, KDD 2025 Datasets and Benchmarks, CIKM 2025 & 2024)，和主席(SIGMOD 2021 Exhibition Chair, DASFAA 2019 Demo Chair)，并多次获得国际会议的重要奖项(SIGMOD 2024 Research Highlight Awards, Best Papers of SIGMOD 2023, VLDB 2021 Distinguished Reviewer Award, ACM SIGMOD 2020 Reproducibility Award, Best Papers of ICDE 2018 & 2012 & 2009, Best Papers of VLDB 2023 & 2015, The Best Paper Award of VLDB 2010)。

## 基于大语言模型和轻量模型结合的多模态数据分析

商烁



**报告摘要：**本研究聚焦Text2SQL技术的突破性训练框架与动态评测基准构建，致力于推动SQL生成模型在数据分析交互过程以及生成准确率的突破。在训练流程方面，本项目使用D2R2-SQL三阶段框架：第一阶段通过多源数据在多维指标下构建高质量训练基座，采用双重模型筛选机制确保数据兼具多样性、难度梯度与可执行性，从而保留模型“可学习临界点”样本；第二阶段引入大小模型协同推理机制，当小模型需要帮助时引入特殊标识符嵌入大模型修正片段，并通过冷启动微调以及以后的强化学习使小模型掌握求助时机，提高推理效率以及准确性；第三阶段为建立执行反馈强化学习闭环系统，在沙箱环境中以可执行SQL、结果匹配度为正向奖励，语法错误/无效执行为负向惩罚，结合数据库实时错误反馈驱动PPO/GRPO算法优化，使模型在自我探索过程中提高SQL能力。针对现有评测系统的不足，本项目设计了A-SQL动态交互基准以弥补现有基准在真实变化场景中的局限性。该基准构建包含模糊查询、晦涩字段名、缺失主外键等现实挑战，要求模型通过多轮交互完成复杂任务。同时在评测维度，引入“自我修正”评价指标，专门设计陷阱任务考察模型解析数据库报错并动态调整SQL的能力，并建立面向过程的评价体系：以任务成功率为核心指标，辅以最终查询准确率验证SQL精确性。通过训练框架与评测基准的双重突破，本研究旨在构建具有模式感知、逻辑推理以及交互能力的Text2SQL系统，为真实复杂SQL场景提供高鲁棒性方法。

**专家介绍：**电子科技大学教授、博士生导师，国家青年特聘专家、国家重点研发计划首席科学家、国家自然科学基金重点类项目负责人。2023、2024 年度国家自然科学基金信息学部会评专家。本科毕业于北京大学，博士毕业于澳大利亚昆士兰大学。研究方向包括大数据、大模型、智能时空计算、智能防灾减灾、社会计算等。在相关领域发表CCF A 类论文 90 余篇，Google Metrics 高影响力/ESI 高被引/热点论文 12 篇，SCI 他引 2000 余次。担任 IEEE TKDE、Geoinformatica、DSE 期刊执行编委，担任 8 个 CCF A 类会议 SIGMOD、VLDB、ICDE、KDD、WWW、AAAI、IJCAI、NeurIPS 的领域主席、（高级）程序委员会委员。研究成果获江西省科技进步一等奖、福建省科技进步一等奖。

## 统一多模态数据融合检索技术

张美慧



**报告摘要：**多模态数据及其价值日益受到企业关注。如何在统一框架下实现高效、准确的多模态数据融合检索，已成为数据管理领域的重要问题。现有技术存在多模态数据表征空间不统一、融合检索效率不足、查询接口复杂等问题，难以满足复杂约束条件下的多模态数据高效融合检索的需求。为此，本项目提出统一的多模态数据融合检索技术框架，包括语义一致的多模态数据表示方法、支持多约束条件的融合索引结构、基于强化学习的多路召回检索机制等。基于以上技术，本项目拟研发多模态数据融合检索原型系统，支持多约束联合查询，提升多模态数据融合检索的性能。

**专家介绍：**北京理工大学计算机学院教授、博士生导师，国家级青年人才。主要研究方向为大数据管理与分析、新型数据管理系统等。近年来，先后主持国家自然科学基金联合基金项目、国家重点研发计划课题等。担任 VLDB 理事会理事、IEEE Transactions on Knowledge and Data Engineering (TKDE) 期刊副主编、VLDB 2024 程序委员会主席，自 2018 年起持续担任 CCF-A 类会议 SIGMOD、VLDB、ICDE 的副主编 (Associate Editor) 或领域主席 (PC Area Chair)，担任 ICDE 2026 Tutorial Track 主席、CIDR 2026-2027 Diversity & Inclusion 主席、EDBT/ICDT 2026 Diversity & Inclusion 主席、EDBT 2022 Demo Track 主席等。荣获 2019 年度计算机学会 CCF-IEEE CS 青年科学家奖、2020 年度 VLDB 青年研究贡献奖，获 VLDB 2019 最佳论文奖、ICDE 2018 及 ICDE 2024 最佳论文亚军、ACM SIGMOD 2019 突出研究奖及 SIGMOD 2024 系统奖等。

## 统一多模态数据融合检索技术

王鑫



**报告摘要：**随着文本、图像、视频、音频等多模态数据呈指数级增长，传统单模态检索系统已难以满足智能问答、个性化推荐及决策支持等前沿应用需求。本项目致力于研发基于统一语义表示的多模态数据融合检索技术，旨在突破现有系统中存在的模态割裂与语义断层等关键瓶颈。研究聚焦三大核心目标：1) 开发多源异构模态数据融合建模与高效索引技术，以提升跨模态检索的效率与精度；2) 设计面向知识图谱与大语言模型深度融合的统一查询接口技术，以实现自然语言智能查询的语义级理解；3) 构建查询执行优化器，以实现复杂多模态查询任务的性能。本研究在知识图谱、大语言模型与多模态数据融合的交叉领域具有显著创新性，其技术成果不仅将填补该领域的研究空白，还将为下一代智能数据管理与检索系统奠定理论基础与技术框架，最终助力我国国产数据库核心技术实现自主可控与跨越式发展。

**专家介绍：**天津大学智能与计算学部教授、博导，人工智能学院副院长。CCF 杰出会员、CCF 信息系统专委会秘书长、CCF 数据库专委会常务委员、CCF 大数据专家委员会执行委员。研究方向：知识工程与大模型、大数据处理。在国内外学术期刊和会议上发表论文 150 余篇。担任多个国际会议程序委员会主席及委员；《计算机工程与应用》副主编、国际期刊 KBS、DSE、WWWJ 编委。曾获天津市科技进步一等奖、中国计算机学会科技进步二等奖，入选高校计算机专业优秀教师奖励计划。

## 统一多模态数据融合检索技术

郑卫国



**报告摘要：**在当前的检索增强生成系统中，知识检索模块的性能直接决定了生成结果的准确性与实用性。“多路召回 + 精排”的组合范式虽然有效缓解了单一检索方式的语义覆盖不足问题，但在实际多模态大数据场景中仍面临诸多挑战：复杂语义融合能力不足、多模态语义对齐与相似性测度挑战、多模态数据孤岛问题严重等。为了解决上述问题，本次报告讲会介绍构建统一的多模态数据融合检索的关键技术，主要包括数据融合建模、图模式约束下的向量检索。

**专家介绍：**复旦大学青年研究员，博士生导师，上海互联网大数据工程技术研究中心副主任。主要从事知识管理、图数据查询和分析、向量计算等相关研究工作，在国际顶级学术会议与期刊发表论文 60 余篇，包括 SIGMOD、VLDB、ICDE、KDD、ACL、NeurIPS、TODS、TKDE、VLDBJ 等 CCF A 类论文，编写知识图谱专著 1 部。担任 Data and Information Management 编委、WISE 2021 Demo chair, GDMA Program Committee Chair, VLDB、KDD、ICDE、IJCAI、TKDE、VLDBJ、TKDD、TOIS 等会议的程序委员会委员和期刊审稿人。

## ■ 14. CCF-阿里云瑶池科研基金论坛

### 14.1 论坛日程

地点：金狮厅会议室

时间：2025年8月2日，星期六，下午14:00~17:00

论坛主席：杜小勇（中国人民大学）、李飞飞（阿里云计算有限公司）

| 时间               | 报告题目                            | 报告人        | 单位        |
|------------------|---------------------------------|------------|-----------|
| 论坛主持人：蔡鹏（华东师范大学） |                                 |            |           |
| 14:00-14:10      | 开场致辞、基金工作整体汇报                   | 伍赛         | 浙江大学      |
| 14:10-14:30      | 面向CXL内存池的近数据计算架构与优化             | 张杰         | 北京大学      |
| 14:30-14:50      | 基于CXL技术的云原生数据库架构设计与优化           | 张一鸣<br>于芯邑 | 厦门大学      |
| 14:50-15:10      | 软硬结合压缩数据直接计算数据库系统研发             | 张峰<br>高诗杰  | 中国人民大学    |
| 15:10-15:30      | 基于GPU和NVMe-SSD的软硬协同向量数据管理关键技术研究 | 柯翔宇        | 浙江大学      |
| 15:30-15:50      | SysInsight基于代码大模型的数据库系统参数调优     | 张心怡        | 中国人民大学    |
| 15:50-16:10      | 面向车联网的海量多模数据压缩技术                | 杨晓春        | 东北大学      |
| 16:10-16:30      | 数据库测试技术研究                       | 姜宇<br>梁杰   | 清华大学      |
| 16:30-16:50      | OLAP的多租任务隔离与运行资源优化研究            | 李环         | 浙江大学      |
| 16:50-17:00      | 论坛总结及致谢                         | 周文超        | 阿里云计算有限公司 |

### 14.2 论坛简介

“CCF-阿里云瑶池科研基金”由阿里云计算有限公司与中国计算机学会（CCF）在2024年联合设立，专注于数据库领域，旨在为领域学者提供科研平台和必要的资源支持，鼓励优秀学者揭榜挂帅，开展与产业相结合的前沿探索性工作，解决真实世界中的卡脖子难题。经技术委员会专家组严格评审，首批8个项目在2024年正式启动。本论坛将邀请基金技术委员会专家组对项目运行进行回顾，并举行8个合作项目的中期报告和研讨活动。本研讨会面向全体NDBC参会者开放，欢迎感兴趣的学者加入！

## 14.3 论坛主席

杜小勇



**专家介绍：** 杜小勇，中国人民大学信息学院教授，博士生导师，中国  
人民大学学术委员会委员，明理书院院长，教育部数据工程与知识工程重点实验室主任，理  
工学部学位评定分委员会主任，信息学院学术委员会主任。数据库课程虚拟教研室和“101 计  
划”数据库系统课程虚拟教研室负责人，CCF 数据库专业委员会资深委员，国家重点研发计划  
项目首席科学家，长期从事数据库及大数据相关领域的研究。主持过国家自然科学基金、国  
家社科基金项目 5 项，国家 973 计划项目、国家 863 计划项目等。曾获国家科学进步二等奖、  
教育部科技进步一等奖，中国计算机学会科学技术一等奖，中国计算机学会卓越服务奖，北  
京市科技进步一等奖等奖项。

李飞飞



**专家介绍：** 李飞飞，阿里巴巴集团副总裁、阿里云数据库产品事业部负责人，ACM、  
CCF 和 IEEE 会士（Fellow）。曾获 ACM 与 IEEE 多项大奖，世界互联网大会 2019 全球  
领先科技成果奖，浙江省科技进步一等奖，中国电子学会科技进步一等奖、中国计算机  
学会科技进步一等奖等。带领团队研发了阿里云企业级云原生数据库系统 PolarDB，云原  
生数据仓库 AnalyticDB，云原生多模数据库 Lindorm，以及云数据库 RDS，连续进入  
Gartner 全球云数据库市场 Cloud DBMS 领导者象限。中国计算机学会（CCF）大数据专  
家委员会常委、数据库专业委员会常委。

## 14.4 论坛主持人

蔡鹏



**专家介绍：** 蔡鹏，华东师范大学数据科学与工程学院教授，博士生导师。2015 年 6  
月加入华东师范大学数据科学与工程学院，在此之前先后就职于 IBM 中国研究院和百度  
(中国)有限公司。在 VLDB、ICDE、SIGIR、ACL 等国际顶级会议上发表多篇学术论  
文。目前的研究方向为内存事务处理、基于机器学习技术的自适应数据管理系统。曾获  
国家科技进步二等奖、教育部科技进步一等奖。

## 14.5 论坛报告

### 开场致辞、基金工作整体汇报

伍赛



**报告摘要：**基金工作整体汇报，2024 年基金运行回顾，2025 年基金发布节奏介绍

**专家介绍：**伍赛，浙江大学计算机学院教授，2011 年毕业于新加坡国立大学博士，主要研究方向：数据库、大数据，数据驱动的智能处理系统。作为主要技术负责人完成批流混合的实时大数据智能处理平台产品研制，成果获得 2016 年教育部科技进步奖一等奖(4/10)与 2019 年中国电子协会科技进步特等奖 (6/15)；参与研发云原生数据库 PolarDB，成果获得 2020 年电子学会科技进步一等奖 (3/15)；参与研发 AI 原生的数据库系统，成果获得 2023 年浙江省科技进步一等奖 (2/13)。 获得 2020 年国家万人计划青年拔尖人才，CCF-A 类会议 VLDB 2014 最佳论文奖，SIGMOD 2023 最佳论文奖。

### 面向CXL内存池的近数据计算架构与优化

张杰



**报告摘要：**本项目的研究目标是利用 CXL (Compute Express Link) 互联协议和 CXL Switch，部分替换现存分布式块存储系统中的 RDMA 互联网络，并在此基础上研究基于 CXL 缓存互联协议的分布式块存储多副本管理机制的优化方案。旨在克服传统分布式块存储系统中多副本管理机制软件开销大、可扩展性差的缺点，设计并开发相关系统原型，满足未来基于 CXL 协议的低延迟、高可扩展分布式块存储系统的要求。

**专家介绍：**张杰，北京大学计算机学院助理教授、博士生导师、特聘研究员。入选国家高层次人才计划海外青年项目，获得英特尔中国学术英才计划荣誉学者、ACM SIGCSE 新星奖。长期从事存储系统和专用处理器的研究和设计，致力于从计算机体系结构层面出发，解决大数据和人工智能时代对于高性能存储系统的需求，突破冯诺依曼体系结构下数据迁移的瓶颈以及内存墙的限制。在国际会议及期刊上发表了五十余篇论文，包括计算机体系结构与系统顶级会议 ISCA、OSDI、HPCA、MICRO、ASPLOS、FAST、ATC、Eurosys 二十余篇。

## 基于CXL技术的云原生数据库架构设计与优化

张一鸣、于芯邑



**报告摘要：**随着互联网技术的飞速发展，CXL 技术的出现给云原生数据库架构的设计和优化带来了新的机会，然而要充分利用 CXL 提供的硬件能力并构建适应 CXL 硬件架构的云原生数据库，面临诸多挑战。本项目围绕索引、并发控制与分布式锁等数据库关键机制和算法展开研究，探索面向三层分离云原生数据库架构的优化设计，主要目标为基于 CXL2.0/3.0 的硬件特性，设计 CXL 友好的数据管理机制和算法，实现高性能可扩展的云原生多主数据库，缩短查询响应时间，提高事务处理速度和内存使用效率。

**专家介绍：**张一鸣，厦门大学教授，厦门市智能存储与计算重点实验室主任。2023 年入选国家级领军人才。长期从事大数据处理技术研究及系统研制。作为第一作者和通讯作者，近五年在 VLDB、ICDE、FAST、EuroSys、ATC、SC 等 CCF-A 类期刊/会议发表论文 27 篇。获 FAST 2023 最佳论文奖、CCF 科技进步一等奖、国家科技进步二等奖、湖南省自然科学一等奖、CCF 优博论文奖等。

## 软硬结合压缩数据直接计算数据库系统研发

张峰、高诗杰



**报告摘要：**近年来，GPU由于其强大的计算能力，在加速数据库系统中的数据处理和查询方面发挥着关键作用，然而有限的GPU内存加上当代大数据的指数级增长给基于GPU的数据库带来了新的困难。为了解决在GPU上高效直接查询级联压缩数据的迫切需求，需要一种专门设计用于增强GPU上直接查询性能的基于Tile的解决方案。本项目旨在将压缩数据直接计算技术开创性地与GPU上的级联压缩方案相结合，充分利用GPU硬件特性，设计一套完整的GPU直接计算算子，并采用更高效的数据布局，实现接近GPU内存带宽的直接查询数据库系统。

**专家介绍：**张峰，中国人民大学教授、博导、CCF 数据库专委委员、高性能专委委员。研究方向为数据库理论与系统，提出压缩数据直接计算理论与技术，围绕该方向以负责人身份先后主持国家自然科学基金、阿里 AIR 等项目，并作为第一或通讯作者发表 CCF A 类论文 40 篇。曾获 TPDS 2021 年度最佳论文、TPDS 2022 Best Paper Runner-up、2024 ASPLOS Distinguished Artifact Awards、ICDE 2024 Industrial and Application Best Paper runner up 等奖励。2022 年入选北京市科技新星人才计划，获得 ACM-SIGHPC 中国新星奖、北京市教学成果一等奖等奖励。任 TPDS 等国际期刊 AE，WISE23 PC Co-Chair、WISE24 Publicity Co-Chair 等，担任 VLDB 等国际会议审稿人。

## 基于GPU和NVMe-SSD的软硬协同向量数据 管理关键技术研究

柯翔宇



**报告摘要：**本项目旨在解决当前海量高维向量数据存储和检索过程中面临的储存消耗高、计算效率低和通信开销大等问题。基于以 GPU 和 NVMe-SSD 为代表的新型硬件，设计 NVMe-SSD 存储特性适配的向量近邻图索引结构、开发 GPU 计算模式契合的索引构建和检索策略、探索 GPUDirect 等 GPU-SSD 直连技术在向量检索上的应用，并将把所提出的模型和算法集成到企业级的阿里云数据库中，以显著提升其处理海量高维向量数据的存储与检索能力，增强其实时计算、分析和决策功能。

**专家介绍：**柯翔宇，浙江大学计算机学院、软件学院研究员，博士生导师。毕业于新加坡南洋理工大学，曾于新加坡南洋理工大学和新加坡国立大学先后从事博士后研究，任 CCF 数据库专委执行委员。在 SIGMOD、VLDB、TKDE 等数据库、数据挖掘领域顶级国际会议和期刊上发表论文 27 篇；获宁波市甬江人才工程科技创新领域青年创新人才项目支持，作为主要参与人参与国家自然科学基金面上项目一项，浙江省“尖兵领雁”重点研发项目一项，宁波市数字经济重点研发专项一项。常年担任 VLDBJ、TKDE 等国内外顶级学术期刊评审专家，KDD、WWW、IJCAI 等顶级国际学术会议程序委员会委员。

## SysInsight：基于代码大模型的数据库系统参数调优

张心怡



**报告摘要：**数据库参数控制着数据库的运行行为，如资源分配、用户连接、查询优化、后台进程等。设置合适的参数值对于数据库系统的性能表现至关重要。本项目提出研发一种能够深入理解数据库系统执行机理的参数调优系统 SysInsight。该系统能够持续低侵入地采样数据库监控指标与系统程序的执行情况，以供分析；并识别高开销的系统函数和程序路径，基于 LLMs 的代码推理给出参数调整建议。SysInsight 具有以下优势特点：深入洞察、精准调优、低成本、可解释性、高泛化性。

**专家介绍：**张心怡，中国人民大学讲师，吴玉章青年英才。博士毕业于北京大学，主要研究方向为自治数据库系统，在该领域已发表 10 余篇 CCF A 类论文，获 2023 年字节跳动奖学金（全球 13 人）、2022 年微软学者提名（亚太地区 21 人）、2019 年 NDBC 优秀学生论文奖等荣誉。研究取得了数据库调优领域的多个突破，包括首个模块化全局调优框架 UniTune（SIGMOD 2023）、首个数据库在线调优系统 OnlineTune（SIGMOD 2021）和首个面向数据库资源优化的元调优系统 ResTune（SIGMOD 2021）。

## 面向车联网的海量多模数据压缩技术

杨晓春

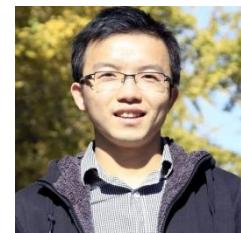


**报告摘要：**智能车联网集成上千个传感器，生成数千至上万的信号指标，其数据类型复杂多样，采集频率也正在从秒级迈向毫秒级，使得企业需要处理的车机信号数据的规模达到 PB 级，车联网中的数据来源于车机、车辆本身、用户、合作伙伴等多个方面，主要分为四种类型：文本、结构化、数值型、视频数据。针对这四种类型数据，本项目研究适用于车机时序数据的高压缩算法，包括：探索信号特征提取，视频数据局部自适应压缩，浮点数据压缩等，压缩效果相比现有算法能有显著提升。

**专家介绍：**杨晓春，东北大学软件学院院长，二级教授，博士生导师。CCF 数据库专委会副主任，CCF 协同计算专委常委。国家级高层次人才，科技部中青年科技创新领军人才，国家优秀青年基金获得者，教育部新世纪优秀人才，辽宁省优秀科技工作者，兴辽计划特聘教授，辽宁省一流课程负责人，沈阳市人大常委会咨询专家；获宝钢优秀教师奖、国家重点研发专项项目首席科学家。《Data Science and Engineering》副主编、《软件学报》编委、《电子学报》编委、《计算机科学与探索》编委。长期从事数据管理与分析领域的教学科研工作，主要研究方向包括大数据管理与知识工程、数据库理论与技术、数据质量管理、数据隐私保护、智能推荐系统等。CCF 杰出会员，ACM 高级会员，IEEE 高级会员。

## 数据库测试技术研究

姜宇、梁杰



**报告摘要：**对数据库系统的可靠性进行测试验证，挖掘可能存在的安全隐患，保障分布式数据库系统的安全可靠是目前学术界和工业界都亟待解决的问题。相比于其他的测试技术，基于变异的模糊测试自动化程度较高且通用性强；基于覆盖率引导的模糊测试引入了数据库系统的覆盖信息作为反馈，比半随机化、全随机化模糊测试输入更具导向性，收敛速度更快，挖掘深层次漏洞更高效，但实际应用中依然面临一些突出难点。在此背景下，本项目研究全面高效的自动化安全测试新方法，突破目前国内关注自动化的数据库系统模糊测试技术工作较少的现状，具有产业上的利用价值。

**专家介绍：**姜宇，清华大学软件学院，长聘副教授。研究方向为软件系统的安全分析，重点关注操作系统、数据库等软件的测试与防护，在 MySQL、Safari 和 Linux 等广泛使用的系统软件中挖掘 300+漏洞被收入中美国家信息安全漏洞库。研发的自动化漏洞挖掘技术被并入谷歌 ClusterFuzz，微软 OneFuzz 等平台。在 ACM SOSP, IEEE S&P, PLDI 等会议和期刊上发表论文 100 余篇，获 EMSOFT, FSE 等会议的最佳论文或提名奖 6 次，主持基金委优秀青年科学基金、科技部重点研发计划项目、华为大颗粒研究计划项目等 20 余项。曾获 2015 年 CCF 优秀博士论文奖、2020 年阿里巴巴达摩院青橙奖，2021 年微众学者，2022 年 NASAC 青年软件创新人才等奖励。

## OLAP的多租任务隔离与运行资源优化研究

李环



**报告摘要：**云数据库供应商面临着资源利用效率与成本效益保证的权衡挑战，严重阻碍其业务规模扩大，该问题在面向更复杂工况和应用需求的无服务数仓场景下更为凸显。本项目旨在探索无服务数仓（Serverless Data Warehouse）环境下的查询调度和动态资源分配问题，以在确保成本效益的同时实现高效资源利用。通过研究新型云数仓实现方式、优化无服务数据库架构，设计安全容错查询调度策略以及弹性资源伸缩技术，为上述云数仓面临的核心难题提供有效解决方案，推动云数仓服务技术进步、提升企业相关产品的竞争力。

**专家介绍：**李环，浙江大学计算机学院研究员，博士生导师。长期从事数据库技术和大数据管理分析方面的研究，其近期工作主要聚焦在新型计算架构（包括云边端、联邦、异构硬件）下查询处理和智能计算的效率优化，成果多发表在数据库和数据智能领域的顶级期刊和会议上，包括三十余篇CCF A类论文，如SIGMOD、VLDB、KDD、WWW、ICDE、ACL、TKDE等；担任多个顶级国际会议的程序委员会委员和知名期刊的客座编辑/审稿人；作为负责人主持国家自然科学基金、国家级创新群体项目课题、浙江省自然科学基金重大项目等，是欧盟“玛丽居里学者”独立研究基金获得者和国家优秀青年基金（海外）获得者。

## 论坛总结及致谢

周文超



**报告摘要：**论坛总结及致谢，2025年基金课题介绍。

**专家介绍：**周文超，阿里云数据库产品事业部总监，AnalyticDB 及生态工具部负责人，负责数据库生态工具与管控的研发以及数据库系统与内核方向科研，专注于Data+AI 场景下的数据库与分布式系统的架构和技术。曾任美国乔治城大学计算机系终身教授，在一流学术会议和期刊发表论文 70 余篇，获 EDBT 和 SIGCOMM 的最佳论文和最佳系统实现奖等奖项，中国计算机协会（CCF）数据库专委会执行委员。

## ■ 15. CCF-蚂蚁科研基金数据库专项论坛

### 15.1 论坛日程

地点：金棕榈厅会议室

时间：2025年8月2日，星期六，下午14:00~17:30

论坛主席：周傲英（华东师范大学）、杨传辉（北京奥星贝斯科技有限公司）

| 时间          | 报告题目                               | 报告人        | 单位                  |
|-------------|------------------------------------|------------|---------------------|
| 14:00-14:20 | 论坛主席致辞                             | 周傲英<br>杨传辉 | 华东师范大学<br>OceanBase |
| 14:20-15:00 | CPU-GPU 协同架构下查询友好的日志压缩算法研究         | 李瑞远        | 重庆大学                |
| 15:00-15:30 | 面向云原生数据库的动态多维度效能感知缓存理论构建与关键技术      | 谢钟乐        | 浙江大学                |
| 15:30-16:00 | 面向超大规模嵌入搜索服务的高性能可扩展向量数据库           | 鲁鹏         | 浙江大学                |
| 16:00-16:30 | 基于向量数据库的多模态视频大模型数据选择与质量优化研究        | 张文涛        | 北京大学                |
| 16:30-17:00 | 知识增强的Text2SQL技术研究：<br>语义理解与逻辑推理的融合 | 李友焕        | 湖南大学                |
| 17:00-17:30 | 基于混合索引的检索增强生成研究                    | 方一向        | 香港中文大学（深圳）          |

### 15.2 论坛简介

数据库内核的高效性是企业运营的基础，云原生数据库和向量数据库则是数据存储与处理的新前沿，而 AI 与 DB 的深度融合正在重新定义数据管理的方式。2025 年发布的 CCF-蚂蚁科研基金数据库专项（第三期），研究方向聚焦于 AI 时代的数据库，主要包括数据库内核、云原生数据库、向量数据库、AI 与数据库融合等产业实际问题。本基金收到广大学者的热情关注和积极申报，在通过 CCF 数据库专委会和蚂蚁集团的专家组成技术委员会的两轮评审后，来自 6 个申请团队的项目获得基金资助，并在 NDBC2025 上进行研究计划的探讨。

## 15.3 论坛主席

周傲英



**专家介绍：**华东师范大学“智能+”研究院院长、贵州（大数据应用）研究院院长、数据学院教授，中国计算机学会会士、常务理事、数据库专委会主任，教育部长江学者特聘教授、国家杰青。主要研究兴趣包括数据库、数据管理、区块链、数字化转型、金融科技等。现担任第八届国务院学科评议组成员、第八届教育部科技委委员、上海市计算机学会理事长、上海市人工智能与社会发展研究会会长、《计算机学报》副主编。曾任复旦大学计算机系主任（1999-2002）、华东师范大学副校长（2016-2023）；获得过国家科技进步二等奖和教育部、上海市技术发明和科技进步一等奖等奖励。

杨传辉



**专家介绍：**杨传辉，现任蚂蚁集团企业级分布式关系数据库 OceanBase CTO。杨传辉曾在百度从事大规模云计算系统研发工作，2010 年作为创始成员之一加入 OceanBase 团队，主导了 OceanBase 历次架构设计和技术研发，从无到有实现 OceanBase 在蚂蚁集团全面落地。同时，他也主导了两次 OceanBase TPC-C 测试并打破世界纪录，著有专著《大规模分布式存储系统：原理与实践》。目前，杨传辉带领 OceanBase 技术团队致力于打造更加开放、灵活、高效、易用的下一代企业级分布式数据库。

## 15.4 论坛报告

### 基于CPU-GPU异构计算架构技术研究

李瑞远



**报告摘要：**当代科技企业每日生成海量的日志数据，这些日志为系统运维、业务分析和系统审计等提供了重要信息。然而，存储海量日志给企业带来了巨大的财务负担。对海量日志的压缩存储能够显著降低企业的成本，降低数据中心的能源使用，助力我国“双碳”目标的推进。现有日志压缩算法存在算法复杂度较高、压缩延迟较大、高效检索困难等问题。本项目拟设计一种 CPU-GPU 协同架构下查询友好的日志压缩算法，提高日志压缩吞吐量的同时，支持海量日志数据的快速无损压缩与高效半解压/不解压查询。

**专家介绍：**李瑞远，重庆大学计算机学院副教授，重庆大学最受学生欢迎的老师，中国电子学会优博，并获得电子信息前沿青年学者出版工程资助。曾经担任京东城市时空数据组机构负责人。加入京东之前，在微软亚洲研究院城市计算组实习 4 年。发表论文 50 余篇（含 CCF A 类 20 余篇，谷歌学术引用 3000+），申请专利 40 余项（含中国专利优秀奖 1 项）。因在时空数据挖掘方面的贡献，获 2024 年 CCF 自然科学三等奖，以及 2025 年空间数据智能新星学者奖。现为 CCF 高级会员、CCF 数据库专委会优秀执委。

## 云原生多维效能感知的智能缓存技术

谢钟乐



**报告摘要：**本项目聚焦于“面向云原生数据库的动态多维度效能感知缓存理论构建与关键技术”，旨在提升云数据库在高并发、混合负载下的性能与稳定性。面对云原生场景中存储异构化、负载时变性及性能指标多目标优化冲突等挑战，项目拟构建统一的缓存效能量化模型，并引入强化学习与时序预测等智能优化机制，实现缓存策略的多维协同与自适应管理。研究将围绕三个核心方向展开：1) 构建融合延迟、迁移开销、一致性代价与负载平衡的效用模型，填补分布式缓存理论空白；2) 设计轻量化强化学习驱动的缓存策略优化方法，支持动态环境下的多目标权衡与在线学习；3) 实现基于访问预测的前瞻性缓存管理机制，有效应对业务高峰和负载突变。本项目计划开发原型系统并最终在 OceanBase 平台上验证其效果，为国产云数据库构建核心技术壁垒、支撑规模化商业化落地提供理论与工程支撑。

**专家介绍：**谢钟乐，浙江大学平台百人研究员，主要研究方向为 AI4System、数据系统构建、ML/DL System 等。在 SIGMOD、VLDB、ICML 等 CCF-A 类会议或期刊发表论文 20 余篇，提交发明专利近 20 项，现/曾担任 VLDB、ICDE 等数据库 CCF-A 类会议程序委员/常驻审稿人，现任 CCF 数据库专委、信息系统专委执行委员，曾获 SIGMOD 2024 System Award、ICDE Best RunnerUp Paper 等荣誉。主持/参与多项国家重点实验室项目、浙江省“尖兵”“领雁”重点研发项目等多项课题。

## 高效大容量向量数据库引擎的研究

鲁鹏



**报告摘要：**近年来，随着大语言模型、计算机视觉和推荐系统的快速发展，由深度学习模型生成的高维嵌入向量已成为图像、音频和文本等复杂非结构化数据的重要表征形式，向量数据库已逐渐成为现代应用的核心基础设施，广泛应用于搜索引擎、推荐系统以及检索增强生成(RAG)等场景。然而，随着数据规模呈指数级增长（单数据集已达 TB 量级）和向量维度持续攀升（如 OpenAI text-embedding-3-large 模型输出维度达 3072 维），传统向量数据库系统在计算开销、动态更新和检索性能、可扩展性等方面都面临严峻的技术挑战。本研究围绕高性能可扩展向量数据库构建中的核心问题，旨在突破现有向量压缩、存储、检索及分布式扩展的瓶颈，旨在构建一个面向超大规模向量数据的高性能检索向量数据库系统。这些突破不仅将推动高维数据表示、近似计算和分布式系统的跨领域理论融合，更为下一代向量数据库的算法设计奠定新的数学基础与架构框架。同时，研究成果将推动国产向量数据库核心技术自主可控，解决 AI 产业化面临的“数据规模爆炸但硬件成本过高”的共性难题，为数字经济提供基础设施级的技术支撑。

**专家介绍：**鲁鹏，博士，浙江大学软件学院平台“百人计划”研究员，高级工程师。2009 年毕业于北京大学，2014 年获新加坡国立大学计算机科学博士学位。曾任微软亚洲研究院系统组访问学者、AT&T 香农实验室访问学者。研究方向为分布式系统、数据库、大规模异构系统软件架构等。近年来，在数据库领域顶级学术期刊 IEEE Transactions on Knowledge and Data Engineering (TKDE) 和国际会议 ICDE、VLDB 等发表论文和报告多篇，获得发明专利 10 余项。先后参与和主持国家重点研发计划、浙江省重点研发计划、浙江省尖兵研发攻关计划、杭州市人工智能重大专项等国家级和省部级科研课题多项，主持企业合作课题 40 余项，获得 2023 年浙江省科学技术进步奖一等奖、浙江省博士后科研项目择优资助(一等)等奖项与荣誉。

## 数据库技术与视频多模态大模型深度融合技术研究

张文涛



**报告摘要：**本项目针对多模态视频大模型训练面临的数据质量参差、检索效率低下等核心挑战，提出基于向量数据库的创新解决方案。研究聚焦视频数据的跨模态特征表示、高效检索机制和自动化质量评估三个关键环节，构建端到端的数据治理框架。通过设计融合时序对齐机制的统一编码器，实现视觉帧、音频流和文本字幕的联合向量化表示；创新性地开发等级相关检索算法，分层建模视频内容中场景、问题和答案的语义关联；同时结合语言模型困惑度分析和视觉-文本一致性检测，建立无参考质量评估体系。该框架突破传统单模态检索的局限性，显著提升跨模态数据检索的精度和效率，并通过动态优化机制持续提升数据质量，为构建高质量视频问答训练数据集提供系统化方法支持。

**专家介绍：**张文涛，北京大学国际机器学习研究中心助理教授、研究员、博导，曾任职于腾讯机器学习平台部、Apple AIML 和加拿大 Mila 人工智能实验室。研究兴趣为大模型数据治理，主持国自然基金青年项目、重大研究计划项目、科技部重点研发计划课题。近 5 年发表 CCF-A 论文 80 余篇，获多个最佳论文奖（如 WWW'22 第一作者，APWeb'23 通讯作者，CIKM'24），领导或参与开源多个机器学习系统。曾获 Apple Scholar、WAIC 云帆奖、北京大学/北京市/中国人工智能学会优博奖、世界互联网大会领先科技成果奖、华为火花奖、中国电子学会科技进步一等奖等荣誉。

## Text2SQL复杂语义理解与推理研究

李友焕



**报告摘要:** 本项目旨在深入研究知识增强的 Text2SQL 技术，致力于融合语义理解与逻辑推理，以解决当前系统在复杂查询场景（如多表关联、嵌套查询、隐含语义及跨领域适应）下的挑战。现有 Text2SQL 模型在数据库模式（Schema）理解、SQL 生成准确率及领域知识适应性方面仍有不足。本研究的价值在于推动“神经—符号协同”框架，提升模型对复杂 Schema 的理解，并加速企业数据分析效率、赋能垂直领域智能化。为实现这一目标，本项目将采用知识图谱增强与神经符号协同的研究方法，构建分层知识图谱并进行 Schema-KG 对齐，设计并应用自动化 Schema 优化算法，基于 Agent 系统和混合架构进行复杂查询分解，并通过适配器模块将领域知识注入模型。本项目拟构建一套通用 Text2SQL 技术体系，推动自然语言数据库查询在数据分析和决策支持领域的产业化应用。

**专家介绍:** 李友焕，湖南大学副教授，博士生导师，CCF 数据库专委会和信息系统专委会执行委员。分别于 2013 和 2018 年获北京大学学士学位和博士学位。2018~2020 年作为腾讯与北大联合培养博士后，在腾讯研究和落地图数据管理和挖掘技术。2021 年 3 月加入湖南大学信息科学与工程学院。目前的主要研究方向基于大模型的数据处理技术。在 SIMOD、VLDB、ICDE、TKDE、IJCAI 等国际顶级会议/期刊上发表二十余篇学术论文，主持国家自然科学基金两项，国家科技重大专项子课题负责人，湖南省自然科学基金一项，授权发明专利三项。

## 基于数据库的检索增强生成研究

方一向



**报告摘要：**近年来，随着生成式人工智能技术的快速发展，大模型已经展现出令人瞩目的能力，并在众多行业中得到了初步应用。尽管如此，大模型在处理特定领域查询问题时仍存在生成内容不正确或“幻觉”、信息过时、缺乏专业领域知识、更新训练代价高等一系列问题。为此，作为应对方案之一的检索增强生成技术（即 RAG 技术）应运而生。RAG 技术通过引入外部知识源来解决大模型在处理特定任务时的知识局限性，增强其在知识获取和生成方面的能力。本项目拟研究基于向量索引、全文索引、标量索引、图索引等多种混合索引的检索增强生成技术，主要包括设计基于图路由的索引选择技术，以及构建基于混合索引查询结果的排序技术，最终提高查询速度和结果准取消，实现数据库与人工智能的深度融合，为用户提供智能化数据查询解决方案。

**专家介绍：**方一向，博士，香港中文大学（深圳）副教授，主要研究面向大规模图数据的数据管理、数据挖掘、检索增强生成、智能体等相关课题，在相关研究领域的高水平会议/期刊上累计发表论文 100 余篇，包括 CCF-A 类论文 70 余篇，第一/通讯作者 CCF-A 论文 50 余篇。其中一项代表性研究成果的论文被评为 SIGMOD 2020 会议的最佳论文之一，并荣获 2021 ACM SIGMOD Research Highlight Award。目前担任 CCF-B 类期刊《Information Processing & Management》的编委、多个顶级会议（如 PVLDB、ICDE、KDD 等）的程序委员会成员、多个顶级期刊（如 TKDE、VLDBJ 等）的审稿人、CCF 数据库专业委员会的执行委员。

## ■ 16. 软件学报专刊论坛

### 16.1 论坛日程

地点：金鸡厅 1 会议室

时间：2025 年 8 月 3 日，星期日，上午 8:30~12:00

论坛主席：李国良（清华大学）、张蓉（华东师范大学）

| 时间          | 报告题目                          | 作者                        | 单位             |
|-------------|-------------------------------|---------------------------|----------------|
| 8:30-8:40   | 论坛主席致辞                        | 李国良<br>张蓉                 | 清华大学<br>华东师范大学 |
| 8:40-9:00   | 基于GPU加速的高维向量聚类算法              | 李忠根,龚盛豪,于浩然,朱轶凡,柳晴,高云君    | 浙江大学           |
| 9:00-9:20   | LSMDiskANN: 一种更新友好型磁盘向量索引框架   | 邱海浪,彭煜玮,彭智勇               | 武汉大学           |
| 9:20-9:40   | 基于大模型的空间数据库自然语言查询转换方法         | 刘孟怡,许建秋,童咏昕               | 南京航空航天大学       |
| 9:40-10:00  | 茶歇                            |                           |                |
| 10:00-10:20 | 权重残差向量量化: 向量压缩与分层索引结构         | 江宇轩,姚俊杰,侯宇轩               | 华东师范大学         |
| 10:20-10:40 | 向量数据库中近似最近邻搜索关键技术综述           | 宋子文,王斌,张喜瑞,赵世豪,杨晓春        | 东北大学           |
| 10:40-11:00 | 嵌入模型微调下的高效 K 近邻图更新方法          | 王嘉翼,徐士惠,李国良               | 清华大学           |
| 11:20-11:40 | GoVector: I/O 高效的高维向量近邻查询缓存策略 | 周依杰,林圣原,巩树凤,余松,范书豪,张岩峰,于戈 | 东北大学           |
| 11:40-12:00 | 面向批量更新的向量索引召回率优化              | 王可,胡思劫,胡卉芪,赵明昊,魏星,屠要峰,周烜  | 华东师范大学         |

### 16.2 论坛简介

本论坛为软件学报专刊“向量数据库及DB4LLM技术”论坛。专刊立足于向量数据管理与分析，探索新型数据库系统核心技术以及向量数据库对支持智能化模型应用的技术突破，特别是聚焦向量数据的高效存取、多模融合的向量建模与计算、向量数据库性能优化、向量数据查询/搜索优化以及其在人工智能技术中的实际应用等方面最新的研究成果。专刊重点关注向量数据库及DB4LLM技术研究中具有创新性和突破性的高水平研究成果，探讨相关基础理论、关键技术，以及在系统研发过程中关于系统设计原理、范式、架构、经验等方面实质性进展，探讨其在相关产业和领域的应用前景以及关键挑战。

## 16.3 论坛主席

李国良



**专家介绍:** 李国良, 清华大学教授, 计算机系副主任, 博士生导师, 国家杰出青年科学基金获得者, IEEE Fellow, openGauss 社区技术委员会主席。曾担任 VLDB Journal、IEEE TKDE 等编委, SIGMOD 2021 大会主席, VLDB 2021 Demo 主席, ICDE 2022 Industry Chair, 多次担任 SIGMOD、VLDB、KDD、ICDE、WWW、IJCAI、AAAI 等会议的程序委员会委员。获得过数据库领域重要国际会议 SIGMOD23、VLDB23、VLDB20、KDD18、ICDE18 最佳论文提名, VLDB23 工业界最佳论文亚军, SIGMOD24 研究亮点奖, DASFAA23 最佳论文、CIKM' 17 的最佳论文奖、DASFAA' 14 的最佳论文提名奖、APWeb' 14 最佳论文奖、EDBT' 13 大数据比赛冠军。获国家科技进步二等奖(2018)、江苏省科技进步一等奖(2019)、电子学会科技进步一等奖(2022)、CCF 科技进步特等奖(2022)。

张蓉



**专家介绍:** 张蓉, 华东师范大学教授, 博士生导师。CCF 数据库专委执委。主要研究方向包括分布式数据管理, 数据库基准评测, 数据库测试、数据流管理等。主持多项国家自然科学基金, 参与多项 863、973 项目。与 PolarDB, OceanBase, 恒生电子, TiDB、交通银行、华为技术有限公司等公司在数据库系统测试、数据库系统 Benchmark 定义、数据管理核心技术研发等领域具有深入合作。已出版英文专著一部、参与翻译《海量数据分析前沿》、《Hadoop 权威指南》(第 2 版和第 3 版);参与编撰教材一本《云原生数据库:原理与实践》。代表性论文发表在 SIGMOD、ATC、ICDE、TKDE、TPDS 等国际顶级会议和期刊。研究成果已经获得国家科学技术进步奖二等奖(第七)、上海市科学技术奖一等奖(第四)。

## ■ 17. 教育论坛

### 17.1 论坛日程

地点：百花厅 2 会议室

时间：2025 年 8 月 3 日，星期日，上午 8:30 ~ 12:00

论坛主席：王鑫（天津大学）、张超（中国人民大学）

| 时间          | 报告题目   | 报告人                            | 单位  |
|-------------|--|--------------------------------|---|
| 8:30-8:40   | 论坛主席致辞   | 王鑫                             | 天津大学  |
| 8:40-9:10   | 中国人民大学数据库系统课程的传承与创新                                | 陈红                             | 中国人民大学                                      |
| 9:10-9:40   | 语言教育科技：智慧教育的新问题                                    | 钱卫宁                            | 华东师范大学                                      |
| 9:40-10:00  | 茶歇   |                                |   |
| 10:00-10:30 | 面向新一代信息技术的山东省计算机科学与技术专业建设：融合与创新                    | 彭朝晖                            | 山东大学  |
| 10:30-11:00 | 计算机学科数据库教育体系建设实践与探索                                | 聂铁铮                            | 东北大学  |
| 11:00-11:30 | OceanBase 产学研人才培养                                  | 徐泉清                            | 蚂蚁技术研究院                                     |
| 11:30-12:00 | Panel讨论：大模型时代下数据库系统<br>课程设计的机遇与挑战<br>主持人：张超 中国人民大学 | 陈红<br>钱卫宁<br>彭朝晖<br>聂铁铮<br>徐泉清 | 中国人民大学<br>华东师范大学<br>山东大学<br>东北大学<br>蚂蚁技术研究院 |

### 17.2 论坛简介

本次论坛旨在探讨新时代下数据库系统课程建设与改革出现的机遇与挑战，聚焦 AI 驱动下的数据库课程体系创新，邀请全国数据库教育专家、行业专家纵谈传承与创新，通过主题报告与圆桌讨论交流分享，深入探讨课程内容重构、教学方法革新、实践体系升级等关键议题，为推动 AI 赋能的数据库教育教学改革建言献策，带动教师精准化教育，促进产学研深度融合，探讨如何培养符合国家战略需求的创新型数据库人才，为新质生产力发展提供坚实支撑。

## 17.3 论坛主席

王鑫



**专家介绍：**王鑫，天津大学英才教授、博导，人工智能学院副院长。国家重点研发计划项目负责人，教育部重点领域教学资源及新型教材建设项目知识工程领域首席专家，教育部重点领域知识工程课程群虚拟教研室带头人。中国计算机学会杰出会员、中国计算机学会信息系统专委会秘书长。研究方向：知识图谱与大模型、知识数据处理。在 IEEE TKDE、SIGMOD、VLDB、ICDE、WWW、IJCAI、《计算机学报》、《软件学报》等国内外学术期刊和会议上发表论文 150 余篇。担任多个国际会议程序委员会主席；《计算机工程与应用》副主编、国际期刊 KBS、DSE、WWWJ 副主编；获天津市教学成果奖特等奖、天津市科技进步一等奖、中国计算机学会科技进步二等奖，入选全国高校计算机专业优秀教师奖励计划。

张超



**专家介绍：**张超，中国人民大学信息学院青年英才、玉兰学者。芬兰赫尔辛基大学博士毕业，清华大学博士后，他作为项目负责人先后承担国家重点研发计划子课题、CCF-华为胡杨林挑战基金项目、CCF-腾讯犀牛鸟项目、达梦产学研项目等。研究方向为数据库系统，相关研究成果在数据库领域期刊/会议上发表论文 20 余篇，曾担任 SIGMOD 2024-2025、ICDE 2023 等数据库领域顶级会议程序委员会委员，他在 HTAP 查询优化方面的研究成果落地 GaussDB 数据库内核，荣获华为火花奖，在云数据库基准的研究工作获得 2025 ICDE Best Paper Runner-Up Award。

## 17.4 论坛报告

### 中国人民大学数据库系统课程的传承与创新

陈红



**报告摘要：**数据库系统是计算机专业的核心课程，最早由老一辈计算机科学家萨师煊教授和王珊教授创建和持续建设，是中国人民大学的传统品牌课程。本报告介绍中国人民大学数据库系统课程教学团队如何紧跟时代需求的变化，围绕教什么、怎么教、怎么学、怎么练、怎么评等教学的基本问题，与时俱进、持续开展教学改革与创新，以更好地服务于新时代人才培养目标。

**专家介绍：**陈红，中国人民大学二级教授，吴玉章讲席教授，数据库与商务智能教育部工程研究中心主任，教育部新世纪优秀人才，国家重点研发计划首席科学家，中国计算机学会数据库专委会常务委员、大数据专家委员会委员、物联网专委会委员。作为项目负责人主持国家重点研发计划项目、国家“核高基”重大科技专项课题、国家863计划项目、国家自然科学基金重点项目等30余项；在KDD、VLDB、ICDE、TKDE等国内外学术期刊和会议上发表论文300余篇。获国家科技进步二等奖、教育部科技进步一等奖和二等奖、北京市高等教育教学成果一等奖、北京市科技进步二等奖、中国计算机学会科技进步一等奖等奖项以及北京市三八红旗奖章、北京市教书育人先锋、北京市优秀教师、首都最美巾帼奋斗者等荣誉称号。主讲的《数据库系统概论》入选国家级线上一流本科课程以及国家级线上线下混合式一流本科课程。学术刊物和会议上发表学术论文70余篇；获山西省科技进步二等奖1项、山西省科技进步一等奖1项、山西省教学成果特等奖2项；主讲的《机器学习》课程被认定为山西省高等学校精品共享课程。博士学位论文获2011年度中国人工智能学会优秀博士学位论文奖。

## 语言教育科技：智慧教育的新问题

钱卫宁



**报告摘要：**经过多年探索，尤其是疫情期间的广泛应用，在线学习已经在不同层次、不同学科的教育领域取得了长足进步。机器语言（编程、SQL）和自然语言（比如中文和英语）教育，虽然属于不同学科，有着截然不同的形式和方法，但是在缺少优质教师和学习资源、难以提供大规模高质量的即时的、个性化的学习服务等方面面临着相同的问题。另一方面，它们又在学习场景、学习服务模式、学习路径构建等方面有各自的挑战。近年来，新兴人工智能技术的发展在为这些问题的解决提供了新工具的同时，也带来了教学模式变革、学习行为建模与理解等一系列新的研究问题。报告将讨论编程和中文在线教育探索中遇到的研究问题和初步结果。

**专家介绍：**钱卫宁，华东师范大学教授、博士生导师，数据科学与工程学院院长，上海市优秀学术带头人，2021 年入选国家级高层次人才计划。目前担任中国计算机学会数据库专委会常务委员，中国计量测试学会计算基准专委会副主任委员，教育部人工智能科技创新专家工作组成员。研究兴趣包括可扩展事务处理，大数据管理系统基准评测，海量数据分析处理及其应用，以及数据驱动的计算教育学。

## 面向新一代信息技术的山东省计算机科学与技术专业建设：融合与创新

彭朝晖



**报告摘要：**为响应山东省高校重点领域教学改革试点项目建设工作（“111 计划”），山东大学携手山东省优势高校和龙头企业，融合教育部“101 计划”建设成果，打造面向新一代信息技术的山东省计算机科学与技术专业发展共同体。本报告将总结“101 计划”成果在山东的推广应用情况，梳理“111 计划”任务目标与建设进展，解析“111 计划”建设思路与实施方案，展示项目创新成果和未来规划，并以“数据库系统”课程为例，分享融合与创新建设的模式与成果。

**专家介绍：**彭朝晖，博士、教授、博士生导师。山东大学计算机科学与技术学院副院长、山东大学（青岛）公共（创新）实验教学中心主任，兼任中国计算机学会数据库专委会执委、大数据专委会执委，山东省人工智能学会高等教育专委会主任，青岛市计算机学会副理事长，山东省计算机“111 计划”工作组秘书长。从事数据库系统、大数据分析、数据治理领域科研与教学工作，主持国家和省部级科研教研项目多项，多次荣获省部级科技进步奖，在国内外知名期刊会议发表学术论文 50 余篇。主持专业核心课“数据库系统”入选首批国家级线下一流本科课程，以第一完成人荣获山东省教学成果奖一等奖，入选高校计算机专业优秀教师奖励计划、青岛高校教学名师，荣获宝钢优秀教师奖、山东大学优秀教师等荣誉称号。

## 计算机学科数据库教育体系建设实践与探索

聂铁铮

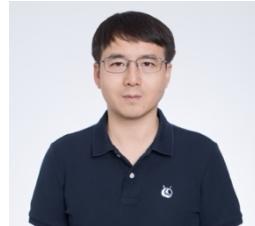


**报告摘要：**数据库原理与技术作为计算机学科的核心内容，是连接计算机理论与实际应用的关键纽带，其课程建设直接影响学生数据思维、复杂系统开发能力及问题解决能力的培养。当前数据库教学面临着理论与实践脱节、教学内容滞后于技术发展、学生学习动力不足等诸多问题。本报告通过介绍东北大学在数据库原理与技术教育的实践经验，探讨面向计算机学科的数据库课程体系、实践平台、创新机制的建设方法与改革方向，包括构建以“基础理论-核心技术-行业应用”为框架的课程体系、以“课程实验-综合实训-科研创新”为体系的实践平台、以“爱国主义-职业道德-科学精神”为目标的思政教育。本报告力求为计算机学科多类型人才培养提供了可借鉴的数据库教育体系建设思路。

**专家介绍：**聂铁铮，东北大学副教授，博士生导师，沈阳工业大学大数据技术创新中心副主任，中国计算机学会高级会员，数据库专委会和信息系统专委会执行委员。长期从事数据管理领域研究工作，研究内容涵盖了数据集成、分布式数据库系统、知识图谱、大数据、区块链系统等多个方向。讲授课程包括“数据库原理”、“数据库管理系统实现技术”、“分布式数据库及应用”、“并行程序设计及应用”和“大数据管理与分析”，教育部“智能基座”数据库虚拟教研室（中国人民大学）成员。承担国家自然科学基金面上项目 2 项、国家重点研发计划子课题 1 项、省部级项目 4 项。承担过 5 项教育部产学合作协同育人项目，长期与国内数据库研发企业合作。曾获得科技部科技进步一等奖 1 项，辽宁省科技进步三等奖 1 项，辽宁省研究生教学成果二等奖 1 项，出版专著和译著 5 部，在国内外期刊或会议上发表学术论文 60 余篇。

## OceanBase 产学研人才培养

徐泉清



**报告摘要：** OceanBase 作为中国分布式数据库领军企业，历经 15 年研发构建原生分布式架构，四次迭代升级至单机分布式一体化架构，两次刷新 TPC-C 世界纪录（7.07 tpmC 超 Oracle 23 倍），支撑“双 11”6100 万次/秒处理峰值，金融行业市占率第一。通过产学研协同育人模式，与 30+ 高校共建联合实验室及课程体系，开发《OceanBase 数据库源码解析》等教材，连续 4 年开展覆盖 500+ 高校、近万人的数据库大赛，累计培养 5.1 万认证人才。技术层面获 500+ 专利、20+ 软件著作权，主导制定 JR/T 0203 等 50+ 技术标准。产品覆盖公有云、多云部署及海外站点，客户突破 1000 家，登榜 Gartner 全球云数据库魔力象限及 IDC 中国分布式数据库厂商评估“领导者”。

**专家介绍：** 徐泉清，北京大学计算机系博士毕业、正高级工程师、蚂蚁技术研究院数据库实验室研究员，目前从事 OceanBase 分布式数据库前沿技术研究工作，浙江省省级海外高层次人才、杭州市全球引才“521”计划高层次人才，研究领域包括分布式数据管理、云数据管理等。曾任新加坡科技研究局(A\*STAR)科学家和新加坡理工大学兼职教师(Adjunct Faculty)，被 A\*STAR 评为研究生导师。先后在国际学术期刊和会议发表 120 多篇学术论文，提交、授权专利 100 多件，担任 VLDB、ICDE 等多个国际学术会议的程序委员会委员以及应邀成为多个国际知名学术期刊的审稿专家，并作为多个标准工作组成员参与相关国际、国家标准的开发与维护。徐博士是国际工程与技术学会会士 (IET Fellow)、CCF 杰出会员和数据库专委会执行委员、ACM 高级会员和 IEEE 高级会员。

## 17.5 Panel

### 大模型时代下数据库系统课程设计的机遇与挑战

陈红、钱卫宁、彭朝晖、聂铁铮、徐泉清

**简介：**大模型不仅重塑了数据存储、处理和分析的技术范式，也为数据库系统课程的教学内容、方法与实践带来了新的机遇与挑战。本 Panel 将围绕以下热点议题展开讨论：（1）大模型驱动的向量数据库、多模态数据处理等新技术，如何融入传统数据库课程？（2）大模型的 SQL 生成与优化能力（如智能查询重写、自然语言交互式 SQL 生成）目前已经很强大，未来课程是否还需要同学花大量精力学习 SQL 语法？（3）未来数据库课程作业应该如何设计才能有效考察出同学的系统能力？

## ■ 18. 贝叶斯数据分析论坛

### 18.1 论坛日程

地点：金熊厅会议室

时间：2025年8月3日，星期日，上午8:30~12:00

论坛主席：岳昆（云南大学）、俞奎（合肥工业大学）

| 时间          | 报告题目                             | 报告人                   | 单位                               |
|-------------|----------------------------------|-----------------------|----------------------------------|
| 8:30-8:40   | 论坛主席致辞                           | 岳昆<br>俞奎              | 云南大学<br>合肥工业大学                   |
| 8:40-9:10   | 贝叶斯决策理论引导的机器学习                   | 曹付元                   | 山西大学                             |
| 9:10-9:40   | 结构信息的智能感知与<br>博弈新理论及方法           | 彭浩                    | 北京航空航天大学                         |
| 9:40-10:00  | 茶歇                               |                       |                                  |
| 10:00-10:30 | 复杂数据偏差下因果推断及<br>赋能大模型思考          | 况琨                    | 浙江大学                             |
| 10:30-11:00 | 信息论视角下的贝叶斯深度学习与<br>图分析技术         | 段亮                    | 云南大学                             |
| 11:00-12:00 | Panel：AI时代的贝叶斯数据分析技术<br>现状、挑战与未来 | 曹付元<br>彭浩<br>况琨<br>段亮 | 山西大学<br>北京航空航天大学<br>浙江大学<br>云南大学 |

### 18.2 论坛简介

贝叶斯定理与算法相结合，催生了众多现代数学定理和值得称道的实践成果，被称为“隐藏在 AI 背后的智能基石”。朴素贝叶斯分类器在文本分类、垃圾邮件检测、情感分析等许多实际问题中取得良好的表现；贝叶斯网用有向无环图描述变量之间的依赖关系、条件概率参数描述其不确定性，成为了不确定知识表示和推理的基本框架，广泛应用于医学诊断、风险管理、故障检测等领域；贝叶斯推理根据新证据不断调整对事件的概率估计，在统计推断、机器学习、人工智能等领域具有广泛应用。大数据、深度学习、大语言模型快速发展并迅速普及，但当代人工智能技术仍面临因果推断能力弱、可解释性差、存在幻觉及被动式响应等问题；而贝叶斯数据分析由于其处理不确定性、因果关系建模、解释复杂现实世界现象等优势，成为了解决上述瓶颈的重要技术手段。本次“贝叶斯数据分析”论坛旨在汇聚国内同行在贝叶斯数据分析领域的最新研究成果，邀请了相关领域专家和青年学者，共同探讨相关领域的前沿技术和研究进展、梳理存在的问题及未来研究方向，进一步促进关键技术创新。

## 18.3 论坛主席

岳昆



**专家介绍：**岳昆，云南大学信息学院教授、博士生导师、院长，云南省智能系统与计算重点实验室主任，研究方向为数据与知识工程。承担国家自然科学基金、云南省基础研究杰出青年项目/重点项目/重大科技专项计划等项目 10 余项。在 TSC、TCYB、TAI、DMKD、AAAI、IJCAI、CIKM、UAI 等期刊及会议发表论文 80 余篇，出版著作 6 部、参编译著 1 部，获云南省自然科学一等奖和青年科技奖。目前任中国人工智能学会理事、因果和不确定性人工智能专委会副主任、人工智能基础专委会委员，中国计算机学会数据库专委会委员，云南省高等学校计算机教指委主任。

俞奎



**专家介绍：**俞奎，合肥工业大学计算机与信息学院教授、博士生导师，研究方向为因果关系推断、因果驱动的可信机器学习。2013-2018 年分别在加拿大和澳大利亚从事博士后研究工作。在 TPAMI、ICML、KDD 等国际期刊与顶级学术会议发表学术论文 80 余篇，出版著作 1 部，获安徽省自然科学奖一等奖、中国计算机学会优秀博士学位论文奖、加拿大太平洋数学科学研究所博士后奖。承担国家自然科学基金、国家新一代人工智能重大专项等项目。目前任安徽省人工智能学会副理事长、中国人工智能学会粒计算与知识发现专委会常务委员、国际期刊 IEEE TETCI 副主编和多个国际顶级会议领域主席。

## 18.4 论坛报告

### 贝叶斯决策理论引导的机器学习

曹付元



**报告摘要：**贝叶斯决策理论，作为一种基于概率推理的决策方法，在机器学习中占据着重要的地位，其核心思想是基于样本的先验概率和类条件概率计算样本的后验概率，然后基于后验概率进行决策。然而，在实际任务中大多难以准确刻画类条件概率，导致贝叶斯决策理论在实际问题中难以发挥作用。本报告从如何有效刻画类条件概率这一问题入手，介绍同质聚簇的基本概念和核心思想，建立同质聚簇与类条件概率的关系，提出可通过同质聚簇的大小估计类条件概率的方法，结合机器学习中的实例选择、标签噪声过滤、贝叶斯错误率估计和模型泛化性能提升等研究问题，介绍相关工作和成果。

**专家介绍：**曹付元，山西大学二级教授，博士生导师，国家高层次人才，计算机与信息技术学院（大数据学院）院长，山西省机器学习与因果推理科技创新人才重点团队负责人，山西省人工智能优秀研究生导师团队负责人，山西省计算机类专业教指委秘书长，山西省人工智能产业技术研究院院长，主要研究方向为机器学习与因果推断。近年来，先后主持国家自然科学联合基金重点项目 1 项、面上项目 3 项、省部级项目 20 余项；在 TPAMI、TKDE、TNNLS、ICML、AAAI、《中国科学》等国际国内重要学术刊物和会议上发表学术论文 70 余篇；获山西省科技进步二等奖 1 项、山西省科技进步一等奖 1 项、山西省教学成果特等奖 2 项；主讲的《机器学习》课程被认定为山西省高等学校精品共享课程。博士学位论文获 2011 年度中国人工智能学会优秀博士学位论文奖。

## 结构信息的智能感知与博弈新理论及方法

彭浩



**报告摘要：**本报告将从结构熵发展新理论、面向网络数据异常检测感知新技术和经典博弈对抗新方法的三个方面做系统介绍。首先将传统简单图的结构熵拓展至多关系图、动态图和双曲图等更逼近现实问题建模的复杂图结构熵新理论；在应用方面，实现了面向开放互联网数据舆情事件检测和异常用户行为检测为核心的可解释、无监督、半监督和有监督的感知新技术，并且在经典单智能体/多智能体/深度强化学习智能博弈对抗中实现了更高效决策新方法，包括在网络水军对抗、强化学习博弈等方面应用，全面提升了强化学习智能决策的性能和自主性。

**专家介绍：**彭浩，北京航空航天大学教授、国家优青。主持国家自然科学基金、国家重点研发计划等科研项目 30 余项。研究领域包括互联网舆情安全、大数据异常检测、强化学习智能博弈等。近五年在 TPAMI、TKDE、TPDS、TC、ACM TOIS、JMLR、AIJ、WWW、NeurIPS、ICML、SIGIR、SIGKDD 等国际旗舰学术期刊或会议发表论文 150 余篇，授权发明专利 40 余件，学术论文获 ESI 热点/高被引论文 11 篇，谷歌学术引用超过 1.2 万次。先后获电子学会技术发明一等奖、中国人工智能学会优秀博士学位论文、仿真学会自然科学二等奖、河北省科技进步一等奖等。

## 复杂数据偏差下因果推断及赋能大模型思考

况琨



**报告摘要：**现阶段以数据驱动、关联学习为模式的机器学习方法（包括深度学习、大语言模型等）倾向于在数据驱动下对变量之间关联关系进行统计建模，缺乏以知识指导机制实现变量之间“由果溯因”的因果推断与分析有效方法。在本次报告中，我将重点介绍因果推断与决策在复杂数据偏差下所遇到的全新挑战，包括（1）未知混杂变量（Unobserved confounders）；（2）选择偏差（collider bias）；（3）同时存在混杂偏差与选择偏差（confounding bias and collider bias）。最后简要介绍从因果回归到因果 transformer 看因果如何赋能大模型的一些思考。

**专家介绍：**况琨，浙江大学计算机学院副教授，博士生导师，人工智能系副主任，主要研究方向为因果人工智能和智慧司法。在 Cell Patterns、TPAMI、ICML、NeurIPS、KDD 等相关领域顶级期刊及会议上发表论文 100 余篇。作为项目/课题负责人承担国家重点研发专项、国家自然科学基金等项目，曾获 ACM SIGAI 中国新星奖，吴文俊人工智能科技进步一等奖，教育部科技进步一等奖，中国电子学会科技进步一等奖。

## 信息论视角下的贝叶斯深度学习与图分析技术

段亮



**报告摘要：**现实世界的大数据是由客观世界生成的，数据之间通常具有复杂的关联关系（即结构），同时也包含噪声和不确定性。数据的知识与规律嵌入在大规模的、包含不确定性的结构中。信息处理的根本任务就是从这些结构中解码出嵌入其中的规律。本报告围绕不确定性知识发现的应用场景，首先介绍基于信息瓶颈与结构信息论的图嵌入方法，以及他们在节点分类、链接预测、时间序列预测方面的应用；然后探讨结合互信息、自编码器和 Transformer 的贝叶斯网嵌入模型，以及他们在贝叶斯网结构学习和概率推理方面的应用。

**专家介绍：**段亮，云南大学信息学院副教授，云南省智能系统与计算重点实验室副主任，云南省兴滇英才支持计划青年人才。研究方向为图数据分析、结构信息论、贝叶斯网学习与推理等。承担国家自然科学基金、中国博士后科学基金、云南省基础研究计划重点/面上项目等项目。在 TKDE、TAI、KAIS、《软件学报》、AAAI、IJCAI、UAI、WSDM、ICDM、CIKM 等国内外期刊及会议上发表学术论文 30 余篇，授权国家发明专利 10 余项，获得云南省自然科学一等奖。

## 18.5 Panel

### AI时代的贝叶斯数据分析技术现状、挑战与未来

曹付元、彭浩、况琨、段亮

**简介：**聚焦当代人工智能前沿技术研究和应用落地，本 Panel 讨论以下热点问题：多模态人工智能背景下的贝叶斯数据分析研究有何潜在创新？AI4DB 背景下的贝叶斯数据分析技术研究主要有哪些开放问题？贝叶斯数据分析与大模型如何双向赋能？

## ■ 19.新一代数据库与数据赋能技术国际论坛

### 19.1 论坛日程

地点：金狮厅会议室

时间：2025年8月3日，星期日，上午8:00~12:00

论坛主席：于明鹤（东北大学）、卢卫（中国人民大学）

| 时间          | 报告题目  | 报告人   | 单位  |
|-------------|---|---|---|
| 8:00-8:10   | 论坛主席致辞  | 于明鹤<br>卢卫                                       | 东北大学<br>中国人民大学  |
| 8:10-8:30   | The Destiny of Database is Data Power Platform  | 周傲英   | 华东师范大学  |
| 8:30-8:50   | AI and Smart Cities: Advanced Technologies for Safe and Comfortable Mobility  | Da Li   | Kyoto Sangyo University   |
| 8:50-9:20   | Big Data Systems for ML   | Kwanghyun Park                                  | Yonsei University   |
| 9:20-9:40   | The Acceleration of Graph Learning based on Ascend NPU  | 张岩峰   | 东北大学  |
| 9:40-10:00  | A Content Recommendation Method Based on Future User Profiles to Support Interest Expansion                                   | Sosui Moribe                                    | Kyushu University   |
| 10:00-10:20 | 茶歇  |   |   |
| 10:20-10:40 | OceanBase: The Fastest Distributed SQL Database for OLTP, OLAP and AI   | 杨传辉   | 北京奥星贝斯科技有限公司  |
| 10:40-11:00 | Industry Intelligence Driven by the Dual Wheels of Data and Knowledge   | 李直旭   | 中国人民大学  |
| 11:00-11:10 | MoE-Enhanced Personalized Federated Learning  | Pengpeng Qiao                                   | Institute of Science Tokyo  |
| 11:10-11:20 | Privacy in Fine-tuning LLM: Attacks, Defenses, and Future Directions  | Yang Cao  | Institute of Science Tokyo  |
| 11:20-12:00 | Panel: The Future of Multimodal Data Management: At the Crossroads of Technological Breakthroughs and Ecosystem Collaboration | 周烜<br>唐博<br>杨世宇<br>丰田正史<br>马强<br>Kwanghyun Park | 华东师范大学<br>南方科技大学<br>广州大学<br>University of Tokyo<br>Kyoto Institute of Technology<br>Yonsei University |

## 19.2 论坛简介

在数据规模爆炸式增长、种类格式日益多样化的时代，大数据与人工智能的快速发展对传统数据库及数据管理方案提出了严峻挑战，尤其在高效处理多源、异构的结构化、半结构化和非结构化数据方面存在明显不足。为突破这些限制，研发面向多模态数据的新一代数据库系统和数据赋能技术至关重要。本次“新一代数据库与数据赋能技术国际论坛”旨在汇聚产学研智慧，探讨统一通用的新一代数据模型、支持柔性解耦与弹性扩展的新型体系结构、跨域数据管理与访问优化技术等前沿技术，促进关键技术创新。本论坛依托于由国家自然科学基金委员会（NSFC）与日本学术振兴会（JSPS）和韩国国家研究基金会（NRF）共同发起设立的合作研究项目“中日韩前瞻计划”（A3 Foresight Program），旨在资助中日韩三国科学家开展合作研究，构建面向未来的数据基础设施，为各行业驾驭复杂数据、释放数据潜能提供核心支撑，驱动数智化变革。

## 19.3 论坛主席

---

于明鹤



**专家介绍：**东北大学软件学院副教授、博士生导师。CCF 高级会员、CCF 数据库专委会、信息系统专委会委员。研究方向为数据库、大数据分析、智慧教育等。承担或参与国家自然科学基金、国家重点研发计划等多项国家级科研项目，及华为、OceanBase 等校企合作项目。在 VLDB、ICDE、TKDE、VLDBJ 等国际重要会议和期刊上发表论文 40 余篇。

---

卢卫



**专家介绍：**中国人民大学教授，信息学院副院长，博导，中国计算机学会数据库专委执行委员，全国大学生计算机系统能力大赛数据库赛道技术负责人。近年来主要从事数据库基础理论、分布式数据库系统等相关领域研究，在 SIGMOD、VLDB 等 CCF A 类会议或期刊发表论文 40 余篇，曾主持国家重点研发计划课题、国家自然科学基金项目、北京市重大专项课题等。主讲的课程《数据库系统概论》获得国家级线上线下混合式一流本科课程和国家级线上一流本科课程。曾获北京市高等教育教学成果一等奖、深圳市科技进步一等奖、首届华为云难题“火花奖”等。

## 19.4 论坛报告

### The Destiny of Database is Data Power Platform

周傲英



**报告摘要:** Before the advent of the internet, databases were widely regarded as the fundamental infrastructure for managing data. Historically, databases and AI, as independent branches of computer science, have intersected multiple times. Contemporary artificial intelligence is essentially data intelligence. Without computers, there would be no internet; without the internet, there would be no big data; and without big data, there would be no current AI. The data intelligence boom is driving a deeper convergence and fusion of databases and AI. In China, data has been defined as the fifth factor of production. Unlike traditional factors of production, it is a driving force akin to electricity, i.e., Data is Power. Therefore, the future direction for databases is to evolve into data power platforms, functioning much like a power grid.

**专家介绍:** ZHOU Aoying, Professor at School of Data Science and Engineering, East China Normal University. He is China Computer Federation (CCF) Fellow and a member of CCF Standing Committee, the Chair of the CCF Technical Committee on Databases, and the Associate Editor-in-Chief of Chinese Journal of Computer. He is the Chairman of Shanghai Computer Society, and the President of Shanghai Society of Artificial Intelligence and Social Development. He was the Chair of the Department of Computer Science, Fudan University (1999-2002), the Vice President of East China Normal University (2016-2023). His research interests lie in Databases, Data Management, Blockchain, Digital Transformation, FinTech, and EduTech.

## AI and Smart Cities: Advanced Technologies for Safe and Comfortable Mobility

Da Li



**报告摘要:** In recent years, the advancement of Mobility as a Service (MaaS) has increased the demand for urban environments that ensure safe and comfortable mobility for all. As part of the broader vision of smart cities, there is a growing need for mobility support technologies that address the needs of vulnerable road users, including the elderly and cyclists. This talk introduces several machine learning-based approaches aimed at enhancing both safety and comfort in urban mobility. Key examples include monocular depth estimation combined with semantic segmentation to detect blind spots, image processing methods to detect objects as reflected in traffic mirrors, and a route recommendation system that estimates bicycle riders' emotions based on upward-facing images of their faces. These studies aim to contribute to the realization of truly inclusive and intelligent urban transportation systems, forming a technological foundation for next-generation smart cities.

**专家介绍:** Da Li (栗達) is an Associate Professor at the Faculty of Information Science and Engineering, Kyoto Sangyo University. He received his Ph.D. in Information Science from Hokkaido University, Japan, in 2020. He served as a research fellow at the Graduate School of Information Science, Hokkaido University in 2020, and as a postdoctoral researcher at Kyoto Sangyo University in 2021. From 2022 to 2025, he was an Assistant Professor at the Faculty of Engineering, Fukuoka University. His research interests include artificial intelligence, natural language processing, machine learning, data mining, recommender systems, and social media analysis.

## Big Data Systems for ML

Kwanghyun Park



**报告摘要:** This talk explores how AI and big data systems are used across industries to derive insights from large-scale data. We focus on Prediction Queries, which involve both data processing and machine learning inference. Traditionally optimized separately, these components present untapped opportunities when jointly optimized. I will introduce Raven, a system that integrates data pipelines and ML workflows for end-to-end optimization, also leveraging hardware accelerators such as GPUs and FPGAs. In addition, I will present Tensor Query Processor (TQP), which transforms SQL queries into tensor programs executable on modern tensor computation runtimes like PyTorch. TQP bridges the gap between database systems and emerging AI hardware/software ecosystems, offering a path for databases to take advantage of rapid innovations in the AI infrastructure space.

**专家介绍:** Kwanghyun Park serves as an Assistant Professor in the Department of Computer Science at Yonsei University, where his primary research interests lie in query processing and optimizations within database management systems. He is also actively engaged in exploring large-scale data processing using innovative hardware architectures, as well as investigating the intersection between data processing and machine learning (ML).

## The Acceleration of Graph Learning based on Ascend NPU

张岩峰



**报告摘要:** Graph machine learning (GML) is widely used in social network analysis, recommendation systems, and bio-computing. However, its computational efficiency faces significant challenges, especially during training and inference on large-scale graph data, which demands highly optimized approaches. In recent years, specialized Neural Processing Units (NPUs) have emerged as a promising solution for accelerating GML tasks due to their efficient parallel computing capabilities. This talk explores how to leverage the powerful computational capacity of the Ascend AI processor—specifically its DaVinci architecture and graph-optimized features—to enhance the performance of GML training and inference.

**专家介绍:** Zhang Yanfeng is a Professor and Doctoral Supervisor at Northeastern University, serving as the Associate Dean of the School of Computer Science and Engineering. He is a recipient of the National High-Level Young Talent Program and the Liaoning Provincial Outstanding Young Scientist Award. His research focuses on database systems and machine learning systems. He has led numerous national research projects, including key initiatives under the National Key R&D Program and the National Natural Science Foundation of China. He has also directed industry collaborations with Huawei 2012 Labs, Alibaba DAMO Academy, and Ant Research. His accolades include the CCF Natural Science Award (Second Class), the ACM SOCC Outstanding Paper Award, and Huawei's Annual Outstanding Project Award.

## A Content Recommendation Method Based on Future User Profiles to Support Interest Expansion

Sosui Moribe



**报告摘要:** One of the primary goals of information recommendation is to help users efficiently discover valuable content. Conventional recommendation methods are well-suited to suggesting content that matches users' existing preferences, but they are not necessarily optimized for expanding users' interests into new domains. This study proposes a recommendation method that supports users in discovering content they can newly enjoy, thereby encouraging the expansion of their interests. Specifically, we generate a "current user profile" that reflects the user's present interests and preferences based on their chat history with an AI agent. Then, the AI agent identifies potential topics or domains that are likely to support interest expansion, and constructs a corresponding "future user profile." From this profile, we extract common topics among content that the user is expected to engage with in the future, and generate search queries based on those topics. The retrieved candidate content is evaluated through a simulated browsing process from the perspective of the future profile, assessing whether each item aligns with the interests represented in the projected user state. Finally, content is recommended alongside suggested perspectives and ways of enjoying it that may help spark new interest and understanding. To validate the proposed method, we focus on music recommendation as a concrete case study.

**专家介绍:** Sosui Moribe is a doctoral student in the Media Design Course at the Graduate School of Design, Kyushu University. He conducts research on interactive AI agent systems and recommendation systems designed to promote user growth. His work focuses on designing dialog-based systems that support personalized learning and development through meaningful interaction. He entered the Department of Art and Information Design, School of Design, Kyushu University in April 2017 and received his B.A. in March 2021. After taking a one-year leave during his master's program, he completed his M.A. in the Content and Creative Design Course at the Graduate School of Design in March 2024. He entered the doctoral program in Media Design in April 2024.

## OceanBase: The Fastest Distributed SQL Database for OLTP, OLAP and AI

杨传辉



**报告摘要:** OceanBase is an open source distributed database with high scalability, high availability and low cost. It started from 2010, and servers over 2000 clients all over the world. It supports hybrid workload, including OLTP, real-time OLAP and AI vector, and sets world records in TPC-C and TPC-H benchmark. Compared to other distributed database, OceanBase is much faster with low cost and low latency. In this presentation, I will introduce OceanBase architecture, technical and product design and some use cases of OceanBase in different scenarios.

**专家介绍:** Charlie Yang is the Co-founder and CTO of OceanBase, Vice President of Ant Group. He joined the OceanBase team in 2010 as one of the founding members. He led OceanBase's product, development and research team, and made OceanBase to be a well-known open-source distributed database. Charlie is determined to lead the OceanBase team to make it more open, flexible, efficient, and easier to use.

## Industry Intelligence Driven by the Dual Wheels of Data and Knowledge

李直旭



**报告摘要:** In the era of artificial intelligence, LLMs have undoubtedly become a key force reshaping the industrial landscape. This speech revolves around the core proposition of industry intelligence driven by the dual wheels of data and knowledge. We first elaborate that the development paradigm of AI has profoundly shifted from "model-centric" to "data-centric", and emphatically highlight the core role of data governance in this transformation. Subsequently, we will examine the "dark side of the moon" of LLMs from a critical perspective, include but not limited to the hallucination problem. Next, we will conduct an in-depth analysis of the unique value of knowledge: how knowledge, as a bridge, forms a synergistic effect with LLMs, and while improving the interpretability of models and enhancing domain adaptability, effectively empowers the implementation of industry-oriented intelligent applications. After that, we will share a series of research results, including the technical framework for the integration of knowledge graphs and large models, innovative methods for injecting domain knowledge, etc. Finally, we will also introduce the practical experience of several university-enterprise cooperation projects, which cover various fields such as manufacturing, finance, and education, vividly demonstrating the remarkable effectiveness of the dual-wheel drive model in solving actual industry pain points.

**专家介绍:** Zhixu Li is a dual-appointed professor and doctoral supervisor at the School of Information and School of Smart Governance, Renmin University of China. He also serves as the director of the Data Engineering and Multimodal Intelligence Laboratory. He has previously held positions as a professor at Fudan University and Soochow University. He obtained his Ph.D. degree in Computer Science from the University of Queensland in 2013. From 2018 to 2021, he served concurrently as the vice president of iFlytek Suzhou Research Institute. His main research directions include LLMs, knowledge graphs, data governance and data engineering, etc. He is a Distinguished Member of the China Computer Federation (CCF), an executive member of CCF Technical Committee on Databases. He has published more than 200 papers in well-known journals and international conferences in the field, and holds over 70 invention patents. He has long maintained industry-university research cooperation with leading AI companies and has rich experience in AI research and practice.

## MoE-Enhanced Personalized Federated Learning

Pengpeng Qiao



**报告摘要：** Federated Learning (FL) enables collaborative model training across decentralized clients, yet effective personalization under data heterogeneity remains a fundamental challenge. This talk presents approaches that leverages model heterogeneity to address data heterogeneity by integrating Mixture-of-Experts (MoE) architectures. For traditional models, we explore personalized federated learning by introducing multiple shared encoders as experts, combined with adaptive client-specific decoders, to enable representation learning aligned with diverse local data distributions. For large language models, we introduce a Sparse-MoE Hybrid PEFT framework that treats each PEFT as an expert, providing a robust and efficient solution for personalization and generalization. This MoE-driven strategy demonstrates how model heterogeneity enables improved personalization, convergence stability, and resource efficiency in federated settings.

**专家介绍：** Pengpeng Qiao is currently a Postdoctoral Researcher at the Institute of Science Tokyo. He received his Ph.D. degree from the Beijing Institute of Technology (BIT) in 2024. His research interests include personalized federated learning, LLM tuning, AI4DB, distributed system, recommendation system, and privacy. He has published at conferences such as ICDE, VLDB, DASFAA, ECMLPKDD, APWeb, GlobeCom, WISE, and at journals such as TKDE, ToN, etc.

## Privacy in Fine-tuning LLM: Attacks, Defenses, and Future Directions

Yang Cao



**报告摘要:** Fine-tuning has emerged as a critical process in leveraging Large Language Models (LLMs) for specific downstream tasks, enabling these models to achieve state-of-the-art performance across various domains. However, the fine-tuning and prompting process often involves sensitive datasets, introducing privacy risks that exploit the unique characteristics of this stage. In this tutorial, I will provide a comprehensive view of privacy challenges associated with fine-tuning and prompting LLMs, highlighting vulnerabilities to various privacy attacks, including membership inference, data extraction, and backdoor attacks. We further review defense mechanisms designed to mitigate privacy risks, such as differential privacy, federated learning, and knowledge unlearning, discussing their effectiveness and limitations in addressing privacy risks and maintaining model utility. By identifying key gaps in existing research, we highlight challenges and propose directions to advance the development of privacy-preserving methods for leveraging LLMs, promoting their responsible use in diverse applications.

**专家介绍:** Yang Cao is an Associate Professor at the Department of Computer Science, Institute of Science Tokyo (Science Tokyo, formerly Tokyo Tech), and directing the Trustworthy Data Science and AI (TDSAI) Lab. He is passionate about studying and teaching on algorithmic trustworthiness in data science and AI. Two of his papers on data privacy were selected as best paper finalists in top-tier conferences IEEE ICDE 2017 and ICME 2020. He was a recipient of the IEEE Computer Society Japan Chapter Young Author Award 2019, Database Society of Japan Kambayashi Young Researcher Award 2021. His research projects were/are supported by JSPS, JST, MSRA, KDDI, LINE, webank, etc.

## 19.5 Panel

---

### The Future of Multimodal Data Management: At the Crossroads of Technological Breakthroughs and Ecosystem Collaboration

周烜、唐博、杨世宇、TOYODA Masashi、MA Qiang、Kwanghyun Park

**简介：**As data ecosystems evolve beyond structured formats, managing multimodal (structured/semi-structured/unstructured) data at scale demands radical innovation. This panel convenes global pioneers to confront core tensions: Can unified data models reconcile generality with performance? How will heterogeneous hardware redefine database architectures? Where do academic breakthroughs intersect industrial realities? We dissect cutting-edge advances in cross-engine optimization, elastic systems, and hardware acceleration while addressing critical ecosystem gaps: interoperability standards, industry-academia translation, and talent development. Join us to chart collaborative pathways for next-generation data infrastructures that turn multimodal complexity into actionable intelligence.

## ■ 20.数据库开源技术论坛

### 20.1 论坛日程

地点：金棕榈厅会议室

时间：2025年8月3日，星期日，上午08:30~12:10

论坛主席：王晨（清华大学）、彭煜玮（武汉大学）

| 时间          | 报告题目  | 报告人                           | 单位                                       |
|-------------|---|-------------------------------|--|
| 08:30-08:40 | 论坛主席致辞  | 王晨<br>彭煜玮                     | 清华大学<br>武汉大学                             |
| 08:40-09:20 | 数据库的开源协同创新发展：以Apache IoTDB为例                                | 王建民                           | 清华大学                                     |
| 09:20-09:50 | 开源开放，生态共赢——OceanBase开源社区技术演进与AI融合实践                         | 徐泉清                           | 蚂蚁技术研究院                                  |
| 09:50-10:00 | 茶歇  |                               |  |
| 10:00-10:30 | A Unified Narrative for Query Processing in Graph Databases | 邹磊                            | 北京大学                                     |
| 10:30-11:00 | AI时代的开源大数据发展现状、挑战和未来  | 李钰                            | 阿里云                                      |
| 11:00-11:30 | 破土·生长·共生：PostgreSQL在中国的演进之路、本土实践与未来展望                       | 张文升                           | PG中文社区                                   |
| 11:30-12:10 | Panel：数据库开源技术在中国的现状、挑战与未来                                   | 王建民<br>徐泉清<br>邹磊<br>李钰<br>张文升 | 清华大学<br>蚂蚁技术研究院<br>北京大学<br>阿里云<br>PG中文社区 |

### 20.2 论坛简介

本论坛将共同探讨开源技术在数据库系统发展中的关键作用。开源模式正在深刻改变数据库技术的研发和应用方式，为学术界和产业界带来新的机遇与挑战。本次论坛将重点讨论以下方向：开源数据库系统的架构设计与优化、开源社区运营与协作经验分享、企业级应用中的技术实践与解决方案，以及开源生态在国内的发展现状与未来趋势。论坛特别邀请国内知名开源项目负责人和资深技术专家进行主题分享，并设置深度交流环节，围绕产学研协同创新、技术自主可控等关键议题展开讨论。通过本次论坛，我们期望促进国内开源数据库技术的交流与合作，推动建立更完善的产学研用协同机制，为学术界提供更开放的研究平台，同时助力产业界实现技术突破。欢迎广大数据库研究者、开发者及开源技术爱好者积极参与，共同推动国内数据库开源生态的繁荣发展。

## 20.3 论坛主席

王晨



**专家介绍：**王晨，副研究员，清华大学大数据系统软件国家工程研究中心总工程师，清华四川能源互联网研究院大数据研究所所长，国家产业基础专家委员会委员。曾任职 IBM 中国研究院主任研究员；参与“数据要素 X”、“可信数据空间”、“十四五大数据产业规划”等国家级数字化战略编制。在数据管理领域三大顶会发表论文 40 余篇，获中美发明专利 60 余项。兼任全国数标委数据治理工作组组长等 20 余项产学研职务，主持/参与制定国家标准十余项。获北京市科技进步一等奖（2020）、日内瓦发明金奖（2022）等荣誉，系 Apache IOTDB 项目 PMC 成员。

彭煜玮



**专家介绍：**彭煜玮，武汉大学计算机学院副教授、博士生导师，研究方向为数据库系统、大数据、数字水印。承担国家重点研发项目、863、自然科学基金重点项目等科研项目，作为技术负责人领导开发了国产对象代理数据管理系统 TOTEM，获教育部科技进步奖一、二等奖各一次。在国内外的著名学术刊物、杂志以及国际著名学术会上发表了数十篇论文。与华为、阿里云、蚂蚁金服、人大金仓、平安科技、北京许继电气等数据库厂商围绕国产自主可控数据库内核研发开展了多次合作。著有《OceanBase 数据库源码分析》、《PostgreSQL 数据库内核分析》。

## 20.4 论坛报告

### 数据库的开源协同创新发展： 以Apache IoTDB为例

王建民

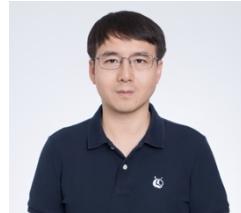


**报告摘要：**开源对数据库产业发展起到了巨大的推进作用。报告以 Apache IoTDB 为例，介绍开源模式下数据库技术创新，及面向 AI 与工业数智化转型时代，数据库技术面临的挑战与趋势。

**专家介绍：**王建民教授，清华大学软件学院院长、大数据系统软件国家工程实验室执行主任。国家杰出青年科学基金获得者。主要研究领域为工业数据软件，包括 CAD 二次开发引擎，产品生命周期管理技术与系统，工业物联网时序数据库管理系统，以及非结构化数据管理技术与系统。以第一完成人获得国家科技进步二等奖、教育部科技进步一等奖、教育部技术发明一等奖、北京市科技进步一等奖，入选北京学者。

## 开源开放，生态共赢——OceanBase开源社区 技术演进与AI融合实践

徐泉清



**报告摘要：**主要围绕 OceanBase 开源社区的技术发展与生态建设成果展开，聚焦“开源开放，生态共赢”的核心理念。通过 1,333 位开发者共建、1,500+企业生产系统落地及社区独立集群数持续增长，OceanBase 以年环比近 400% 的速度从边缘走向核心，构建起多云融合的分布式数据库生态。系统梳理了 OceanBase 对 MySQL 生态的深度致敬与能力补全，通过分布式架构、实时数仓、向量检索等技术实现“新一代 MySQL”的定位。同时，OceanBase 全面拥抱 AI，推出智能开发、向量数据库、MCP 插件等创新方案，融合 Spring AI Alibaba 生态，助力 Data x AI 场景落地。未来，OceanBase 将持续推进开源协作，拓展 AI 时代数据库边界，打造存储与智能一体化的“新一代 MySQL”。

**专家介绍：**徐泉清，北京大学计算机系博士毕业、正高级工程师、蚂蚁技术研究院数据库实验室研究员，目前从事 OceanBase 分布式数据库前沿技术研究工作，浙江省省级海外高层次人才、杭州市全球引才“521”计划高层次人才，研究领域包括分布式数据管理、云数据管理等。曾任新加坡科技研究局(A\*STAR)科学家和新加坡理工大学兼职教师(Adjunct Faculty)，被 A\*STAR 评为研究生导师。先后在国际学术期刊和会议发表 120 多篇学术论文，提交、授权专利 100 多件，担任 VLDB、ICDE 等多个国际学术会议的程序委员会委员以及应邀成为多个国际知名学术期刊的审稿专家，并作为多个标准工作组成员参与相关国际、国家标准的开发与维护。徐博士是国际工程与技术学会会士（IET Fellow）、CCF 杰出会员和数据库专委会执行委员、ACM 高级会员和 IEEE 高级会员。

---

## A Unified Narrative for Query Processing in Graph Databases

邹磊



**报告摘要:** With the advent of graph data, graph databases have garnered significant research interest and efforts in recent years, especially with respect to graph query processing. There have been a vast suite of methods for efficient graph query processing, especially for the core graph query constructs, regular path queries (RPQs) and subgraph matching queries (SMQs). In the meantime, there is an observable divide among these methods as well as confusion between them and their relational counterparts. In this talk, we provide a unified narrative for graph query processing, so as to bridge the gap between existent lines of work and offer a comprehensive view of the query processing workflow in graph databases.

**专家介绍:** 邹磊，北京大学王选计算机所教授，大数据分析与应用技术国家工程实验室（北京大学）知识集成和智能决策中心主任。邹磊教授的研究包括图数据库，知识图谱，尤其是基于图的知识图谱数据管理，面向知识图谱的自然语言问答，图分析与机器学习，以及大数据系统；他已经发表了 50 余篇计算机领域的顶级国际学术期刊和会议（CCF-A）上，包括 SIGMOD, VLDB, ICDE, TODS, TKDE, VLDB Journal 等。邹磊教授曾获得中国自然科学基金（NSFC）优秀青年基金，国家重点研发计划项目的资助。另外，还获得了英国皇家学会的牛顿高级访问学者基金的资助。2017 年获得教育部自然科学二等奖（获奖项目“大规模图结构数据管理”，排名第一）。

## AI时代的开源大数据发展现状、挑战和未来

李钰



**报告摘要：**在 AI 时代的大背景下，数据的存储和访问模式以及数据生成和访问的主体都在发生变化，这对数据平台架构提出了新的挑战和需求。报告将全面介绍开源大数据架构的现状、面临的挑战、演进的趋势和相应的解决方案。

**专家介绍：**李钰，阿里云 EMR 负责人，阿里巴巴资深技术专家，ASF Member, Apache Flink&HBase&Paimon PMC Member，具备多年阿里巴巴集团内部和阿里云上开源大数据计算&存储大规模生产实践经验。

## 破土·生长·共生：PostgreSQL在中国的演进之路、本土实践与未来展望

张文升



**报告摘要：**分享 PostgreSQL 中国开源社区的发展轨迹与公益实践。聚焦社区自发组织、志愿贡献、知识沉淀的演进历程，剖析用户互助、人才培养、本地化推广的核心经验。探讨社区治理、国际协作、技术普惠面临的挑战与机遇。展望构建更包容、活跃、可持续的社区生态，促进纯粹技术交流与全球开源共治。

**专家介绍：**张文升，PostgreSQL 中国社区负责人，天工开物基金会 TOC，《PostgreSQL 实战》《PostgreSQL 指南：内幕探索》等书作译者。近年聚焦于数据库国产化与人工智能应用。

## 20.5 Panel

---

### 数据库开源技术在中国的现状、挑战与未来

王建民、徐泉清、邹磊、李钰、张文升

**简介：**本次 Panel 将聚焦三大核心议题：首先探讨中国数据库开源生态的当前格局，包括主流技术路线选择、产学研协作模式及典型应用场景；其次分析关键技术挑战，如开源社区可持续运营机制、核心研发人才梯队建设、企业开源战略与商业化的平衡等现实问题；最后展望未来发展方向，重点讨论 AI 驱动的数据库自动化运维、信创背景下自主可控技术路径、以及开源标准体系建设等前沿趋势。讨论还将特别关注中国开发者如何在全球开源生态中构建差异化竞争力，以及开源协议演进对技术自主性的影响等热点话题。

## ■ 21. 移动云大云海山数据库合作生态论坛

### 21.1 论坛日程

地点：奥斯卡厅会议室

时间：2025年8月3日，星期日，上午08:30~11:50

论坛主席：伍赛（浙江大学）、丁岩（中移（苏州）软件技术有限公司）

| 时间          | 报告题目   | 报告人      | 单位                         |
|-------------|--|----------|----------------------------|
| 8:30-8:40   | 论坛主席致辞   | 伍赛<br>丁岩 | 浙江大学<br>中移（苏州）软件<br>技术有限公司 |
| 8:40-9:10   | 时空数据库自然语言查询转换方法<br>研究                            | 许建秋      | 南京航空航天大学                   |
| 9:10-9:40   | 面向云原生数据库的Serverless数据<br>压缩传输技术研究                | 顾荣       | 南京大学                       |
| 9:40-10:10  | 大云海山数据库产品与架构解析                                   | 丁岩       | 中移（苏州）软件<br>技术有限公司         |
| 10:10-10:20 | 茶歇   |          |                            |
| 10:20-10:50 | 面向新兴计算架构的高效图数据处<br>理系统研究                         | 孙世轩      | 上海交通大学                     |
| 10:50-11:20 | 如何在K8s上构建大规模的DBaaS<br>(Database as a Service) 平台 | 曹伟       | 杭州云猿生数据库<br>有限公司           |
| 11:20-11:50 | 关键行业数据库云化的思考和实践                                  | 吕伟初      | 金纂信科有限责<br>任公司             |

### 21.2 论坛简介

本论坛以“学术筑基·生态共创”为核心，汇聚顶尖高校学者与行业技术领袖，深度探讨数据库领域的前沿研究、系统创新与生态协同。论坛依托移动云“大云海山”技术体系，聚焦云原生架构、时序数据库和图数据库等方向，搭建产学研高效对话桥梁，推动国产数据库核心技术突破与生态繁荣。

## 21.3 论坛主席

伍赛



**专家介绍:** 伍赛, 浙江大学计算机学院教授, 2011 年毕业于新加坡国立大学博士, 主要研究方向: 数据库、大数据, 数据驱动的智能处理系统。作为主要技术负责人完成批流混合的实时大数据智能处理平台产品研制, 成果获得 2016 年教育部科技进步奖一等奖(4/10)与 2019 年中国电子协会科技进步特等奖(6/15); 参与研发云原生数据库 PolarDB, 成果获得 2020 年电子学会科技进步一等奖(3/15); 参与研发 AI 原生的数据库系统, 成果获得 2023 年浙江省科技进步一等奖(2/13)。获得 2020 年国家万人计划青年拔尖人才, CCF-A 类会议 VLDB 2014 最佳论文奖, SIGMOD 2023 最佳论文奖。

丁岩



**专家介绍:** 丁岩, 移动云海山数据库总师&高级技术研究专家, 三十多年软件研发工作经验, 主持研发电信资源管理信息系统、3G 核心网网管、智能网平台、IPTV 管理平台、WoStore、大数据平台、分布式数据库等多个大型软件产品。正高级工程师, CCF 数据库专委会执行委员, 计算机行业协会数据库专委会常务理事, ACMUG (MySQL 中国用户组) 主席团成员 & 南京分会主席, 东南大学计算机学院硕士生企业导师, 南京航空航天大学人工智能学院硕士生行业导师, 南京邮电大学通信与信息工程学院企业导师, 四十多件授权发明专利、论文、奖项, 参与《云原生数据库技术》标准编写。曾任中兴通讯 GoldenDB 创始人&研发负责人&首席科学家、平凯星辰 TiDB 数据库首席科学家。

## 21.4 论坛报告

### 时空数据库自然语言查询转换方法研究

许建秋



**报告摘要:** 近年来, NL2SQL 技术发展迅速, 其为非专业用户提供了友好便捷的任务输入方法。该技术不仅在数据管理领域, 而且在数据分析、软件运维等领域都将有着广泛的应用价值。本报告主要聚焦时空数据这一垂直领域, 主要从以下三个方面就课题组近期的进展进行介绍: (1)基于大模型的空间数据库自然语言查询转换方法, (2)高质量空间数据库自然语言查询语料库构建, (3)基于大模型的移动对象自然语言查询转换基准测试。同时, 探讨未来继续深入研究的方向, 包括通用语料检测和修复、自然语言查询转换系统的综合评估等。

**专家介绍:** 许建秋, 南京航空航天大学教授, 计算机系主任, ACID 研究所所长, 主要从事数据管理及数据库原型系统研究, 近年来主要针对大模型驱动的复杂数据管理和高效计算、垂直领域智能数据管理和分析开展工作, 主持国家自然科学基金项目 3 项、省部级项目 4 项以及 CCF 华为胡杨林-数据库专项基金。发表学术论文 50 余篇, 包括中国计算机学会推荐 A 类论文和中科院 1 区论文等, 例如 IEEE TKDE、ICDE、PVLDB、计算机学报等, 出版英文专著 1 本。担任中国科学卓越期刊二期领军、中科院 2 区期刊 *Frontier of Computer Science* 青年编委, 多次获得国际/国内学术会议论文奖, 包括最佳系统演示论文奖, 最佳展望论文奖等, 授权国家发明专利 7 项。担任多个数据库领域国际/国内优秀学术期刊评审专家如 TODS、TKDE、计算机学报、软件学报(2024 年度优秀审稿人)等, 也是多个 CCF 推荐 A 类国际会议的程序委员会委员如 PVLDB、ICDE、KDD 等。

## 面向云原生数据库的Serverless数据压缩传输技术研究

顾荣



**报告摘要：**云原生数据库由于其交付便捷、极致弹性、更低成本等优势，近年来得到广泛使用。然而，云上分布式跨域存储的特点，给数据库中很多相关数据传输操作（例如：备份、同步）造成严重性能瓶颈。本报告中，我将介绍一种利用 Serverless 技术实现云原生数据库数据经济高效传输的方法。其背后的难点在于平衡云上数据传输的性能与成本影响因素复杂。我们首先对 Serverless 场景下的数据库数据传输影响因素进行分析建模。然后，针对单任务、多任务以及动静态场景，分别提出云上数据压缩传输的精准高效参数搜索方法。主流公有云平台上的实验结果表明，本方法在速度成本上优势明显。本报告工作发表 CCF A 类论文多篇，并在实际系统中发挥较好成效。

**专家介绍：**顾荣，南京大学特聘研究员/博导，国家级青年人才计划入选者，CCF 南京秘书长，CCF 杰出演讲者，研究方向为大数据智能计算系统、云计算系统，研究成果发表于 SIGMOD, VLDB, USENIX ATC, EuroSys, VLDBJ 等国际顶级会议期刊。他曾获得达摩院青橙奖（2023）、江苏省科学技术一等奖、IEEE TCSC Early Career Excellence Award、HPCC 会议最佳论文奖等科研学术荣誉，以及华为火花奖、阿里优秀学术合作项目奖、中兴产学研优秀项目奖等业界合作奖项。他牵头研制云原生弹性数据缓存系统 Fluid 入选国际云原生计算基金会（CNCF），广泛应用于阿里云、腾讯云、火山引擎、天翼云、360、小米等行业领军型企业。他被选为 Fluid 开源社区主席，持续推动相关技术创新发展。

## 大云海山数据库产品与架构解析

丁岩



**报告摘要：**移动云自研大云海山数据库如何助力企业数字化转型？本报告深入剖析海山数据库在一体化、计算与存储深度融合、HTAP 混合负载、行列混存与多模数据管理等方面的核心技术与工程实践。

**专家介绍：**丁岩，移动云海山数据库总师&高级技术研究专家，三十多年软件研发工作经验，主持研发电信资源管理信息系统、3G 核心网网管、智能网平台、IPTV 管理平台、WoStore、大数据平台、分布式数据库等多个大型软件产品。正高级工程师，CCF 数据库专委会执行委员，计算机行业协会数据库专委会常务理事，ACMUG（MySQL 中国用户组）主席团成员 & 南京分会主席，东南大学计算机学院硕士生企业导师，南京航空航天大学人工智能学院硕士生行业导师，南京邮电大学通信与信息工程学院企业导师，四十多件授权发明专利、论文、奖项，参与《云原生数据库技术》标准编写。曾任中兴通讯 GoldenDB 创始人&研发负责人&首席科学家、平凯星辰 TiDB 数据库首席科学家。

## 面向新兴计算架构的高效图数据处理系统研究

孙世轩



**报告摘要：**图作为建模和分析实体关系的重要手段，已广泛应用于社交网络、在线支付、互联网等实际场景。然而，图数据本身的海量性、稀疏性和异构性，以及图处理过程中的数据动态性、负载动态性和计算动态性，给大规模图计算中的性能优化与硬件资源高效利用带来了严峻挑战。为应对这些问题，我们聚焦于面向多核 CPU、GPU 和 Serverless 等新兴计算架构的图数据处理研究，结合图数据与计算负载特性，采用算法与系统协同设计的方法，充分挖掘底层架构的性能潜力，提升整体系统效率。本次报告将重点介绍我们在两方面的最新进展：一是面向多核 CPU 高并发场景的动态图内存存储技术；二是面向 Serverless 架构的弹性图计算系统设计。

**专家介绍：**孙世轩，博士，现任上海交通大学计算机科学与工程系长聘教轨副教授。2020 年于香港科技大学获得博士学位，2020 年至 2023 年于新加坡国立大学从事博士后研究员工作，2023 年入职上海交通大学工作至今。申请人长期从事数据库系统和并行计算等方面的研究，入选了国家级高层次青年人才计划和上海市海外高层次青年人才计划，博士期间获得香港政府奖学金 HKPFS。近年来专注于高性能图数据库系统和面向新兴计算架构的图计算加速方面的研究。在 SIGMOD、VLDB、ICDE、ASPLOS 等数据库/体系结构 CCF-A 类刊物发表论文二十余篇。受邀担任了 VLDB'24-25、ICDE'24-26、SC'25、EuroSys'26 等数据库、高性能计算和系统领域 CCF-A 类会议的程序委员会委员，长期担任数据库三大 CCF-A 类期刊 TODS、VLDBJ、TKDE 的审稿人。

## 如何在K8s上构建大规模的DBaaS (Database as a Service) 平台

曹伟



**报告摘要:** 在云原生技术飞速发展的今天，如何在 Kubernetes 上构建一个大规模的 DBaaS (Database as a Service) 平台，已经成为技术创新的核心挑战之一。而 KubeBlocks，作为一款开源工具，为这一目标提供了极具吸引力的解决方案。它通过标准化的数据库定义模板和自动化的运维能力，让在 K8s 上集成一款数据库像搭建乐高积木一样简单灵活。KubeBlocks 支持 MySQL、PostgreSQL、Redis 等 40 多种数据库，能够在 Kubernetes 集群中实现快速部署、弹性扩展和高可用管理，确保数据的可靠性和一致性。此外，它还提供多租户隔离支持，结合统一的 API 接口，帮助企业构建安全、高效的云原生数据库服务平台。可以说，KubeBlocks 将传统复杂的数据库运维转化为一种标准的、可操作性的工程实践，为大规模 DBaaS 平台的构建提供了强大的工具与坚实的基础。

**专家介绍:** 曹伟，杭州云猿生数据库有限公司创始人兼 CEO。前阿里云数据库总经理/研究员，云原生数据库 PolarDB 发起人。中国计算机学会数据库专委会执行专委、开源专委会执行专委，获得 2020 年中国电子学会科技进步一等奖，在 SIGMOD、VLDB、ICDE、FAST、USENIX ATC 等数据库与存储国际顶级学术会议发表论文 20 余篇。

## 关键行业数据库云化的思考和实践

吕伟初



**报告摘要:** 聚焦金融、通信等关键行业，分析其数据库云化的核心需求与独特挑战，探讨可行的云化路径和关键技术思考，并分享行业在分布式架构、云原生改造等方面的探索实践，旨在平衡创新与稳健。

**专家介绍:** 吕伟初，金纂信科有限责任公司总工程师，江苏省产业教授，信创工委会 WG2 数据库产业推动组组长，北京金融科技产业联盟特聘专家，基础软件质量控制与技术评价工信部重点实验室专家委员。获得江苏省、通信学会等科技进步奖 4 项，获得授权专利 12 项，参与《安全可靠分布式事务型数据库技术要求》等多项重要标准的制定。

## 22. 萨师煊优秀学生论文汇报

### 22.1 报告日程

报告地点：金鸡厅 2 会议室

报告时间：2025 年 8 月 3 日，星期日，上午 8:30~12:00

| 时间          | 报告题目   | 作者   |
|-------------|--|--|
| 8:30-8:40   | WookongDB查询重写技术研究  | 徐言诗、张心怡、卢卫、杜小勇   |
| 8:40-8:50   | A Pilot Study of Text-to-SQL Parsing for Traditional Mongolian                 | 邢瑞俊、王少鹏  |
| 8:50-9:00   | DoPI: Doctor-like Proactive Interrogation LLM for Traditional Chinese Medicine | Zewen Sun, Ruoxiang Huang, Jiahe Feng, Rundong Kong, Yuqian Wang, Hengyu Liu, Ziqi Gong, Yuyuan Qin, Yingxue Wang, Yu Wang |
| 9:00-9:10   | 基于图神经网络的数据高效基数估计模型   | 张书赫、康辉、周施羽、张力璇、王新颖   |
| 9:10-9:20   | 高效视频文本和视频问答数据质量评估方法  | 梁昊、陈梓榕、崔斌、张文涛  |
| 9:20-9:30   | 集成多推理范式的表格问答方法研究   | 吴桐宇、赵广振、孔德畅、董振江  |
| 9:30-9:40   | 基于自适应少样本提示的检索增强生成查询优化方法  | 崔奕、周明君、曾维新、赵翔  |
| 9:40-9:50   | SieveJoin: Boosting Multi-Way Joins by Filtering Unneeded Intermediate Results | 李人瑞、马清志  |
| 9:50-10:00  | 基于混合内存架构的动态哈希索引优化  | 刘高聪、袁以规、金培权  |
| 10:00-10:10 | DynaPart: 面向近似最近邻搜索的自适应分区索引框架  | 史鹏程、杨世宇、杨建业  |

## 23. 分组报告

### 23.1 分组报告1：数据智能与分析

报告地点：奥斯卡厅会议室

报告时间：2025年8月2日，星期六，下午14:00~17:30

| 时间          | 报告题目   | 作者                 |
|-------------|--|--------------------|
| 14:00-14:06 | 融合对抗混合负采样的图对比学习推荐算法  | 宋威、王田靖、宁可庆、郭威      |
| 14:06-14:12 | AGES: Multimodal Data Processing for Scene Recognition   | Hong Daocheng      |
| 14:12-14:18 | 融合知识关系演化的动态图知识追踪模型   | 吴邦振、闫秋艳、王儒增        |
| 14:18-14:24 | MMBRel: 面向中文编程学科知识的层次化实体关系抽取模型   | 田秀霞、李冰雪、裴壮         |
| 14:24-14:30 | Agent4Stu: 基于大语言模型的学生作答行为高效模拟智能体   | 刘苏熠、刘淇、高维博         |
| 14:30-14:36 | UTCANet: 时间卷积与注意力门控协同驱动的时序预测模型   | 陈裕中、杨喆、金福生         |
| 14:36-14:42 | 基于多视图跨模态特征增强与融合的图像描述生成   | 张乃洲、曹薇             |
| 14:42-14:48 | 动态二分图剪裁的 HBase 两阶段多副本亲和算法  | 施华峰、杨伟伟、陈勃胜        |
| 14:48-14:54 | 表格型文本数据中的关系抽取  | 薛成杰、周玉玲、黄泽昊、岳昆、王笳辉 |
| 14:54-15:00 | 基于张量图扩散的共享近邻密度峰值聚类算法   | 王健、赵可钦             |
| 15:00-15:06 | 智能知识发现中主体自觅行为的特质刻画   | 高金胜                |
| 15:06-15:12 | Multi-Scale Contrastive Graph Reconstruction-Based Heterophilic Graph Representation Learning              | 孙驰、潘海为、张可佳、江海英     |
| 15:12-15:30 | 茶歇   |                    |
| 15:30-15:36 | FNG: Graph Algorithms for Approximate Nearest Neighbor Search with Structured and Unstructured Constraints | 孟祥磊、聂铁铮、申德荣、寇月     |
| 15:36-15:42 | 多模态时序知识图谱实体对齐研究  | 周仲憬、赵润豪、张时琪、曾维新、赵翔 |
| 15:42-15:48 | 针对基于图神经网络谣言检测器的可复用攻击模型   | 范业博、李逸成、刘勇、张薇      |
| 15:48-15:54 | 基于图编辑距离的表达性嵌入与差异建模   | 丁辉湖、李瑞滔、张伟超、刘钰     |

| 时间          | 报告题目                    | 作者                |
|-------------|-------------------------|-------------------|
| 15:54-16:00 | 带有多域去噪的属性感知序列推荐         | 周品超、王楠、王润哲、李金宝    |
| 16:00-16:06 | 基于核心实体与局部子图的知识图谱增量更新方法  | 彭商濂、娄颖、冯丽         |
| 16:06-16:12 | 融合审计制度知识图谱的大语言模型微调及推理方法 | 吕鹏、于天宇、孙孜、郭信好、林培光 |
| 16:12-16:18 | 基于大模型的低人力科研知识图谱构建       | 陈昱鹏、严沐阳、姚俊杰       |
| 16:18-16:24 | 对比学习与标签动态增强的短文本多意图检测    | 向贵良、陈昌奉、张扬英       |
| 16:24-16:30 | 双重监督驱动的图对比聚类学习研究        | 谭思婧、彭磊、张聪、卢虹妃     |
| 16:30-16:36 | 互信息客户端聚类的分组异步联邦学习       | 李方博、王楠            |

## 23.2 分组报告2：数据系统与应用

报告地点：百花厅 1 会议室

报告时间：2025 年 8 月 3 日，星期日，上午 8:30-12:00

| 时间          | 报告题目  | 作者   |
|-------------|---|--|
| 8:30-8:36   | 融合先验知识与贝叶斯优化的高鲁棒性数据库参数调优系统  | 陈小强、陈小强、孙辛远、李博涵、屠要峰  |
| 8:36-8:42   | A Multi-aspect Label Information Fusion Enhanced Method For Automatic ICD Coding      | Yonghe Yan, Yujing Xin, Ning Liu, Qian Li, Dongbiao He, Teng Ma, Lei Liu, Zhongmin Yan |
| 8:42-8:48   | 基于改进生成对抗网络的缺失数据恢复方法   | 张素、邓松、刘庆生  |
| 8:48-8:54   | 大模型驱动的空间数据自然语言查询语料库构建方法研究   | 易炜佳、刘孟怡、许建秋  |
| 8:54-9:00   | MTune：蒙特卡洛树搜索在数据库管理系统配置调优中的应用   | 高锦涛、胡志远、姜璐璐  |
| 9:00-9:06   | 基于结果复用的视频目标检测   | 李忠斌、李欣辉、刘斌、岳昆  |
| 9:06-9:12   | 基于物理算子图的自适应基数估计方法   | 高锦涛、余晓婷  |
| 9:12-9:18   | 一种具有可解释性的司法典型案例筛选方案   | 卫晓妍、张柏礼  |
| 9:18-9:24   | OLAP 数据库性能的 CPU 关键硬件影响因素研究  | 王沁垚、韩瑞琛、刘佳茹、张延松  |
| 9:24-9:30   | 面向关系型数据库锁机制实现的模糊测试框架  | 杨洪宇、翁思扬、陆辰昊、张蓉、周烜  |
| 9:30-9:36   | 大容量高并发下数据库稳定性优化研究   | 胡谱赟、潘巍、马泽琪、韩宇翊、沈阳、李战怀  |
| 9:36-9:42   | 自动视频理解提示词生成系统   | 梁昊、张文涛   |
| 9:42-9:48   | 知识图谱引导的发电量预测  | 薛成杰、向冰冰、岳昆、王笳辉   |
| 9:48-9:54   | Predicting Agricultural Ecosystems: A Dynamic Food-Web Modeling Approach              | 陈奕蒙、石雅宁、沈子钧、张文涛  |
| 9:54-10:00  | LIGHT: A Learned Join Order Selection Network Using Dynamic Graph Attention Mechanism | 张秋鸿、杨建文、严瑾、丁治明、张硕  |
| 10:00-10:06 | 电力系统中不平衡样本的数据增强方法   | 邓松、王亮  |
| 10:06-10:12 | 基于双延时深度策略梯度的时序数据库调优方法   | 黄震乾、叶枫、吴峥、余书磊、李娜   |
| 10:12-10:30 | 茶歇  |  |
| 10:30-10:36 | 面向小型边缘设备的缓存算法比较研究   | 陈雨欣、丁好、程建勋、柴云鹏、张怡龙、孙国道、梁荣华、刘健  |

| 时间          | 报告题目  | 作者                |
|-------------|---|-------------------|
| 10:36-10:42 | Hierarchical Multi-Modal Rumor Detection via CNN-Transformer Fusion and Co-Adaptive Cross-Modal Integration | 李逸成、范业博、刘勇、张薇     |
| 10:42-10:48 | A Self-Iterative Framework on Text2Cypher Based on GPT-4o   | 牛放、王朝坤            |
| 10:48-10:54 | 基于多特征融合的列车司机疲劳驾驶检测  | 王海涌、田爱爱、张丹        |
| 10:54-11:00 | Workload-Aware Buffer Prefetching for Database Systems  | 易杰、金培权            |
| 11:00-11:06 | AD-SQL:融合审计提示和分步策略的 SQL 查询优化模型  | 郭信好、林培光、孙攷、于天宇、吕鹏 |
| 11:08-11:12 | 面向海量电商用户的异构社交图构建及应用   | 张子谦、王朝坤、冯昊、吴呈     |

## 24. 系统演示

### 24.1 论坛日程

地点：长廊

时间：2025年8月3日，星期日，上午8:30~12:00

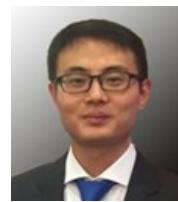
论坛主席：徐辰（华东师范大学）、赵翔（国防科技大学）

| 论文题目  | 作者   |
|---|--|
| 基于RAG的评论数据观点问答系统  | 邱江涛  |
| KDSelector: A Knowledge-Enhanced and Data-Efficient Model Selector Learning Framework for Time Series Anomaly Detection | Zhiyu Liang, Dongrui Cai, Chenyuan Zhang, Zheng Liang, Chen Liang, Bo Zheng, Shi Qiu, Jin Wang, Hongzhi Wang |
| 高效可信任的大规模合成数据生成系统   | 许琪、叶哲名、盛鼎、毛晓一、张召、金澈清   |
| SCorpus: 高质量空间数据自然语言查询语料库构建工具   | 易炜佳、黄唯、许建秋   |
| SAGE: A Dataflow-Native Framework for Modular, Controllable, and Transparent LLM-Augmented Reasoning                    | 刘沛林、张书豪、刘俊、张森磊、张睿诚   |
| NeuroMem: A Programmable Memory Framework for Building Personalized LLM Assistants                                      | 张书豪、张睿诚、王笑涵  |
| 轻量级跨链单据存证管理系统   | 张文字、钱惠勋、李仲恒、刘雨石、张召、金澈清   |
| POQD: Performance-Oriented Query Decomposer for Multi-vector retrieval  | 刘耀阳、李俊霖、吴垠鋆、陈震   |
| Artemis: A Customizable Workload Generation Toolkit for Benchmarking Cardinality Estimation                             | 陆辰昊、胡梓锐、张蓉、杨章程、周烜、徐泉清、杨传辉  |
| 基于模型上下文协议和多数据库的<br>负荷预测系统设计与实现  | 徐瑞翔、黄福兴、叶锐、袁培森   |
| Uniplore: 大语言模型驱动的自动数据科学平台  | 许玉田、李晖、王延昊、夏圣杰、闵圣天   |
| TGraphDB: 面向属性值演化的时态图数据管理系统   | 左宗宇、孙照海、宋景和、林学练、马帅   |
| Graph Association Analyses for Early Drug Discovery<br>(特邀VLDB 2024最佳演示系统论文)  | Wenfei Fan, Daji Li, Peiyu Liang, Shuhao Liu, Yaoshu Wang, Yiming Wang, Min Xie, and Runjie Zhang            |
| scRAG: an Efficient Retrieval Augmented Generation System for scRNA-seq Data Analysis<br>(特邀ICDE 2025最佳演示系统论文)          | Yuren Mao, Yifan Zhu, Qing Liu, Peigen Liu, Yu Haoran, Yunjun Gao  |

| 论文题目   | 作者  |
|--|---|
| IsoVista: Black-box Checking Database Isolation Guarantees (特邀演示系统论文)                                      | Long Gu, Si Liu, Tiancheng Xing, Hengfeng Wei, Yuxing Chen, David A Basin   |
| UniView: A Unified Autonomous Materialized View Management System for Various Databases (特邀演示系统论文)         | Zhenrong Xu, Pengfei Wang , Guoze Xue, Qitong Yan, Shenghao Gong, Yelan Jiang, Yuren Mao, Yunjun Gao, Shu Shen, Wei Zhang, Dan Luo, Lu Chen |
| DocDB: A Database for Unstructured Document Analysis (特邀演示系统论文)  | Zequn Li, Yuanhao Zhong, Chengliang Chai, Zhaoze Sun, Yuhao Deng, Ye Yuan, Guoren Wang, Lei Cao   |
| VQFT: A Visual Query Approach Based on Full-Text Search for Knowledge Graphs (特邀演示系统论文)                    | Zhaozhuo Li, Xin Wang, Meng Wang, Yajun Yang, Bohan Li, Dong Han  |
| Andromeda: Debugging Database Performance Issues with Retrieval-Augmented Large Language Models (特邀演示系统论文) | Pengyi Wang, Sibei Chen, Ju Fan, Bin Wu, Nan Tang, Jian Tan   |
| NeutronRAG: Towards Understanding the Effectiveness of RAG from a Data Retrieval Perspective (特邀演示系统论文)    | Peizheng Li, Chaoyi Chen, Hao Yuan, Zhenbo Fu, Hang Shen, Xinbo Yang, Qiange Wang, Xin Ai, Yanfeng Zhang, Yingyou Wen, Ge Yu                |
| PixelsDB: Serverless and NL-Aided Data Analytics with Flexible Service Levels and Prices (特邀演示系统论文)        | Haoqiong Bian, Dongyang Geng, Haoyang Li, Yunpeng Chai, Anastasia Ailamaki  |
| IKGA: An Interactive Visualization Tool for Knowledge Graph Alignment (特邀演示系统论文)                           | Weixin Zeng, Shiqi Zhang, Huang Peng, Zhen Tan, Weidong Xiao, Xiang Zhao  |

## 24.2 论坛主席

徐辰



**专家介绍:** 徐辰, 华东师范大学数据科学与工程学院教授、博士生导师。2014 年-2018 年担任德国柏林工业大学高级研究助理, 参与 Apache Flink 系统的研发。2014 年获华东师范大学博士学位, 曾于 2011 年赴澳大利亚昆士兰大学短期访问。主要研究兴趣是大规模分布式数据处理系统, 相关研究成果发表在 PPoPP、SIGMOD、VLDB、ICDE、TKDE 等 CCF-A 类国际会议和期刊, 出版 Springer Briefs 学术专著一部, 编著《分布式计算系统》教材。入选上海市青年科技英才扬帆计划, 主持中央军委科技项目、国家自然科学基金面上和青年项目、上海市自然科学基金面上项目、CCF-腾讯犀牛鸟基金项目等, 参与 2 项国家重点研发计划项目和 1 项国家自然科学基金重大研究计划项目。作为主要完成人获得 2020 年上海市科技进步一等奖、2022 年上海市优秀教学成果一等奖、2022 年国家级教学成果二等奖。

赵翔



**专家介绍:** 赵翔, 国防科技大学大数据与决策实验室智能决策技术研究部主任, 教授、博导。教育部青年长江学者, 湖南省科技创新领军人才, 军队青年科技英才。从事大数据知识工程、世界模型等方面的教研工作, 主持国家重点研发计划青年科学家项目、国家自然科学面上基金、湖南省杰出青年基金等重要科研项目 10 余项, 出版专著 2 部, 获评国际学术会议论文奖 4 项, 申请专利 40 余项。湖南省科技创新团队骨干成员, 获评军队科技进步二等奖、中国指挥控制学会科技进步一等奖、中国计算机学会自然科学二等奖、湖南省高等教育教学成果一等奖、湖南省教学科学研究优秀成果二等奖。中国计算机学会杰出会员、数据库专委执行委员。

## 25. 赞助商风采

### 25.1 赞助单位

长期钻石合作伙伴



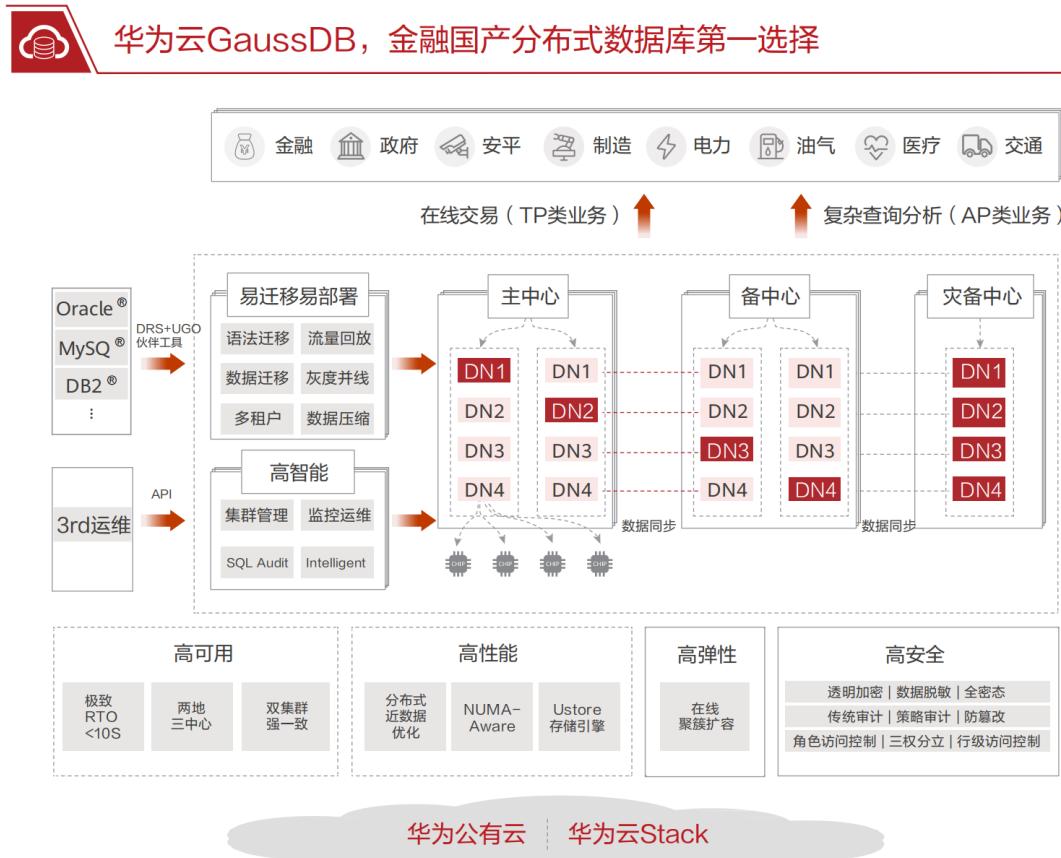
长期铂金合作单位



长期金牌合作单位



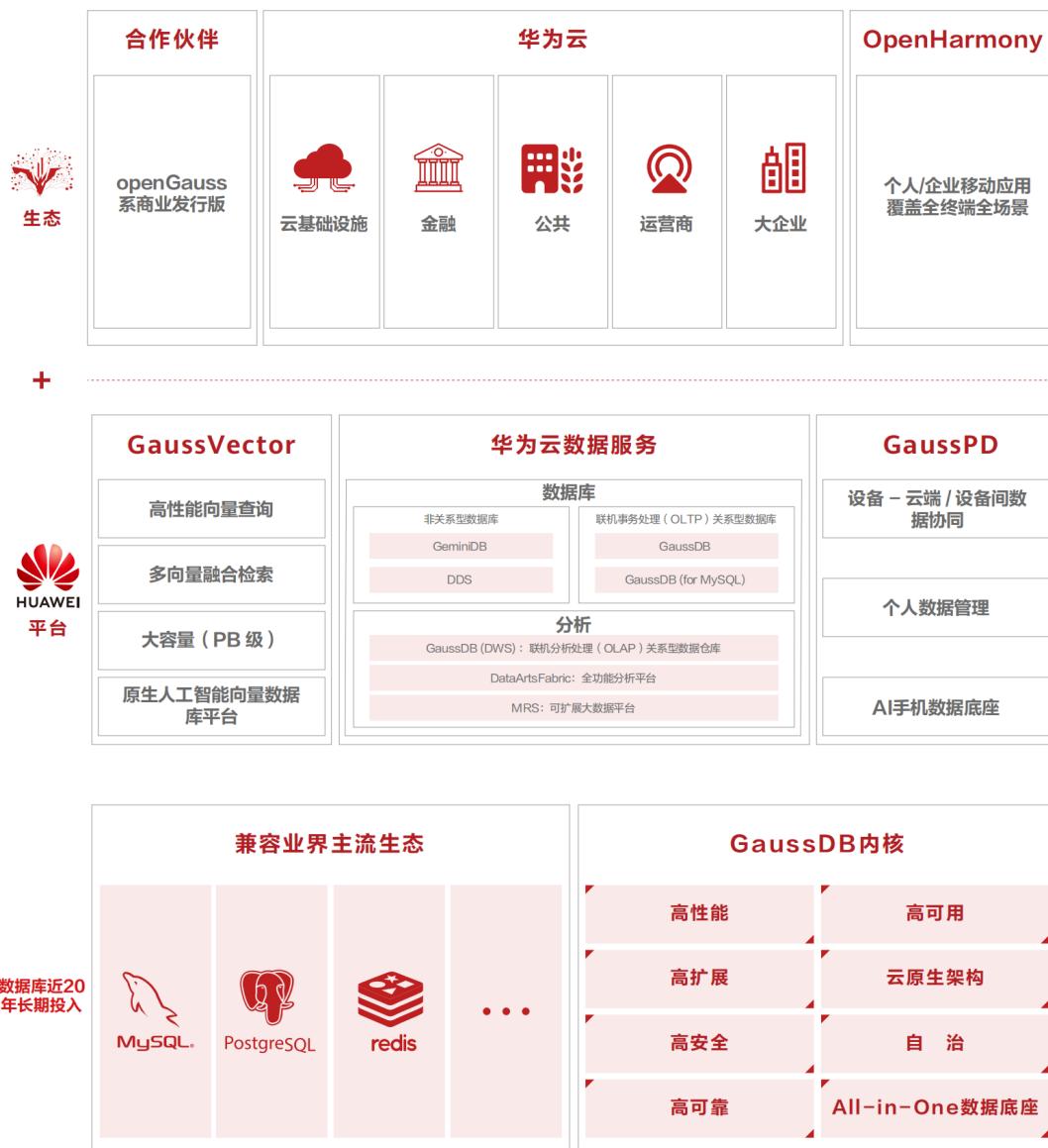
## 25.2 长期钻石合作单位：华为



### 易部署

Oracle数据库、MySQL数据库常用语法高度兼容，结构+数据一站式迁移解决方案，让不确定变确定

## 华为数据库：持续基础研究投入，开源开放生态，打造领先数据库





# Gauss松鼠会

## 关于我们

Gauss松鼠会是一个由华为公司数据库团队发起、联合高校/企业及志愿者共同运营的数据库技术交流圈。

Gauss松鼠会汇集数据库的爱好者和关注者，大家共同学习、探索、分享数据库前沿知识和技术，像松鼠一样剥开科学的坚果；交流Gauss及其他数据库的使用心得和经验，互助解决问题，共建数据库技术交流圈。



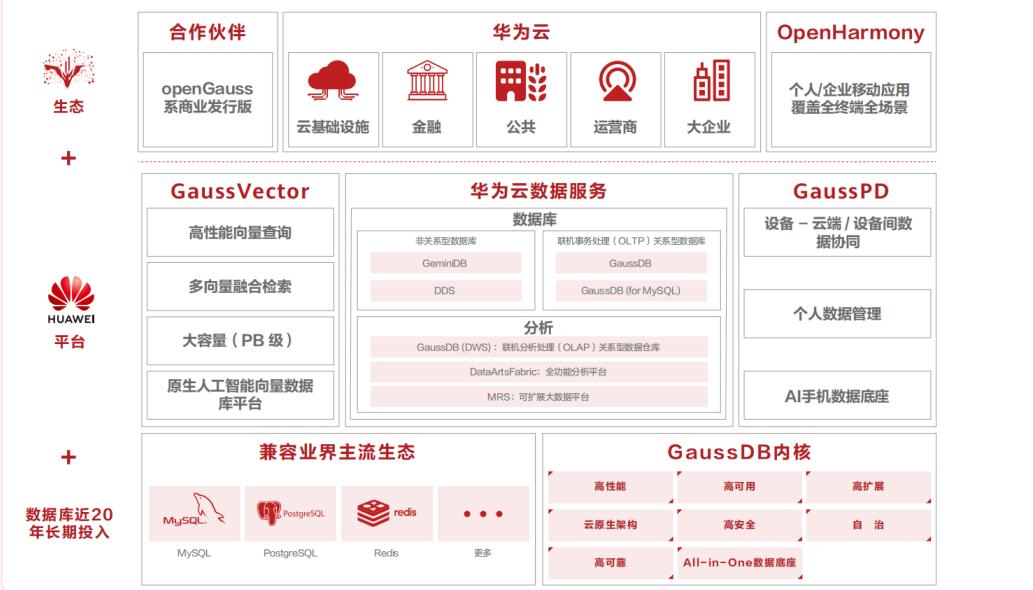
Gauss松鼠会公众号



Gauss松鼠会小助手



### 华为数据库：持续基础研究投入，开源开放生态，打造领先数据库



### 华为2012实验室中央软件院高斯部（招聘）

高斯数据库二十年深耕数据库内核技术，形成了可以满足通讯、IT、云、工业控制等多场景的系列数据库产品，涵盖了云原生数据库、分布式数据库、嵌入式数据库、图数据库等核心技术，与众多合作伙伴、科研机构、数据库专家等一起创建了openGauss企业级数据库社区，产学研用协同创新共发展，旨在立足面向全球构筑一个全新的数据库生态，和全行业一起打造数字全场景数据库。

#### 数据库架构师

负责数据库内核架构、设计及竞争力关键技术研究，构建数据库内核能力。  
 1、参与学术和客户数据管理趋势和技术发展，识别数据库内核技术竞争力；  
 2、结合云、硬件、网络和数据库理论的技术发展，面向客户服务发展，完成关键技术构建。

#### 数据库产品资料开发工程师/布道师

1、负责数据库产品运维资料（例如日常管理、故障处理、性能调优等）或开发者资料的设计、开发和维护，确保资料高质量交付；  
 2、通过友商分析、用户调研、可用性测试等方式，识别产品资料改进点，持续提升产品资料可用性和易用性。

#### 数据库内核工程师

打造业界领先的数据库核心技术，面向未来云化、分布式、跨平台、人工智能等高新技术场景构建高性能、低时延、高安全、高可靠的下一代数据库技术并实现商业成功。  
 1、负责数据库系统模块设计，包括但不限于查询引擎、执行引擎、存储引擎等特性设计开发，性能调优等工作；  
 2、基于存储、计算、网络、调度等技术构建数据库高性能、高可用，软硬件协同等核心竞争力；  
 3、负责数据库相关领域的新技术探索和关键技术突破；  
 4、参与下一代数据库的设计、开发与原型验证。

#### 数据库测试工程师

1、对数据库的产品交付质量和产品特性进行评估和验收；  
 2、洞察数据库领域的先进测试理论、方法和工具，持续牵引团队测试能力提升。



联系人：刘女士  
 电 话：18611628368

## 25.3 长期钻石合作单位：达梦数据

达梦数据 股票代码 688692

全栈数据产品与解决方案提供商  
自主研发原始创新 构建信息安全基石 打造网信核心力量

中国电子旗下  
网信领域战略核心企业

中国数据库标准委员会组长单位

国家企业技术中心

国家技术创新示范企业

博士后科研工作站

### 达梦数据 国产之光

2000年，达梦数据成立。作为国产数据库领军企业，达梦始终坚持自主研发，拥有完全自主知识产权，掌握了数据库领域的核心技术，构建了中国数据库的“代码根”。公司核心团队在数据库领域拥有40余年研发经验及技术积累，为大中型企业、事业单位及政府提供各类数据库软件及集群软件、云计算与大数据产品、数据库一体机等一系列数据库产品及相关技术服务。

依托自主研发技术优势，坚持以客户为导向，达梦产品市占率连续多年位居中国数据库管理系统市场国内数据库厂商前列。如今，达梦产品已广泛应用于关乎国计民生的金融、党政、央国企、电力、能源、通信、交通、医疗等多个关键行业，产品装机量达数百万套，服务于数亿人的工作及生活。

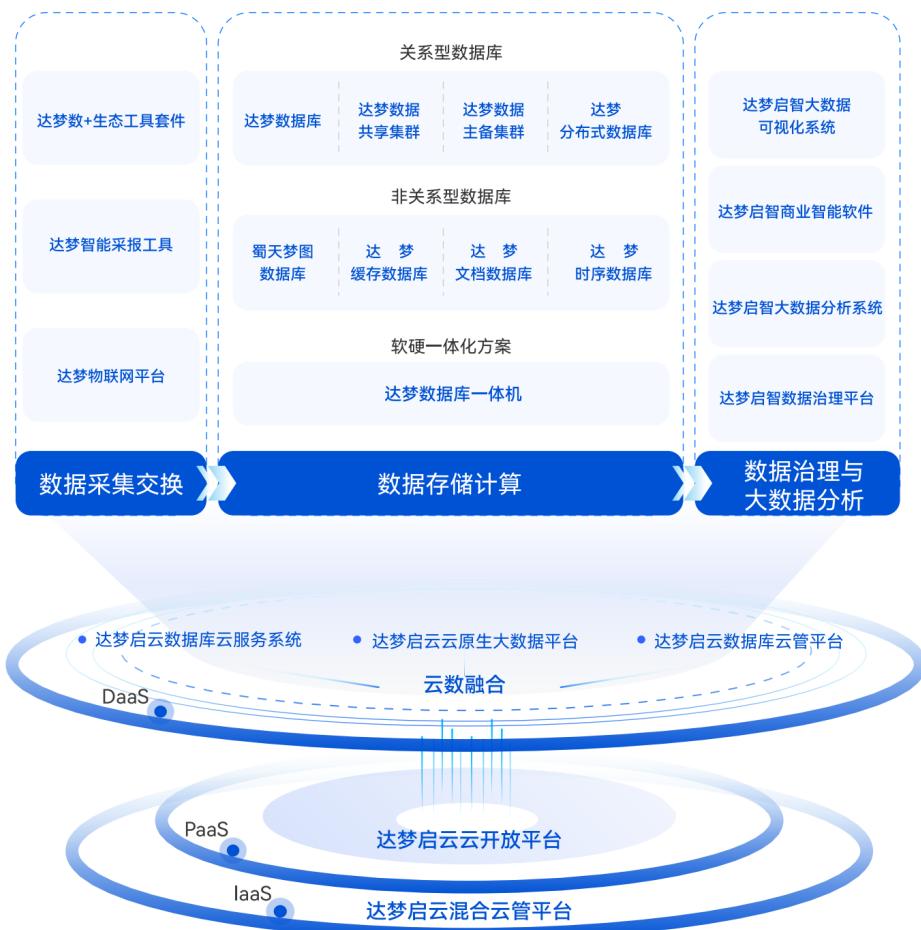
累计代码  
**60000000 行** 成功构建  
**中国数据库的代码根**

**99.99%** | **40余年** | **遥遥领先** | **60+项**

自主原创 数据库领域深耕 国产数据库市场占有率 国家及省部级荣誉资质

## 自主为根 技术为核

达梦以自主为根、以技术为核，立志掌握中国人自己的核心技术。通过组建强有力的研发团队，加大研发投入，不断加快技术创新，推出了以达梦数据库为代表的核心数据库产品，并掌握了数据库技术的皇冠——共享存储集群核心技术，拥有产品全部核心源代码，筑牢我国数字信息安全基石。



一键扫码  
获取更多产品资料



## 生态共建 合作共赢

达梦坚持生态共建、合作共赢，为合作伙伴赋能。与重点行业千余家头部ISV及渠道伙伴达成深度合作，共建自主生态。目前，达梦与产业链上下游多家厂商完成12000余款产品兼容适配，通过培训、赋能活动等多样化形式，为生态伙伴持续赋能，实现共建、共享、共创、共赢的数据库产业生态。

13000+ 2000+

产品兼容适配

签约合作伙伴

达梦联合多家软硬件厂商进行技术攻关，构建产业生态



## 产教融合 人才强企

达梦作为高技术人才密集型企业，技术人员占比高达70%。深入贯彻落实人才强国战略，达梦高度重视人才，积极协同产业链、创新链、教育链、人才链优势培育人才。



### 博士后科研工作站

为达梦插上“创新之翼”

2022年，达梦获批国家级博士后科研工作站，加大自主创新投入，引入博士后科研人员，培养高层次科研人才，促进产学研融合，提高企业核心竞争力。以博士后科研工作站为抓手，达梦加强人才梯队培养。推进核心研发人员培养，助力产品及技术创新。加大技术服务、运维人员培养，致力于为客户创造价值。



### 人才培养体系

为达梦注入“动力之源”

达梦积极推动数据库人才培养及考核，基于人才职业发展生命周期，推出层次化的人才培养体系，包括达梦认证培训体系、1+X证书体系及定制化培训体系。



### 人才解决方案

为达梦引入“创新之魂”

达梦“岗课赛证”综合育人解决方案，以人才培养为核心，以深化产教融合为宗旨。目前，达梦已与上百所院校建立合作，举办了软件杯、工匠杯、达梦杯等多个赛事，培养了行业人才数万人，并成功获批“湖北省产教融合型企业”，推动产学研用深度融合。



一键扫码  
参与合作调研

## 25.4 长期铂金合作单位: OceanBase



OceanBase 是 100% 根自研的原生分布式数据库，基于单机分布式一体化架构，高度兼容 Oracle 和 MySQL，支持事务处理 (TP)、实时分析处理 (AP) 和 AI 多工作负载融合，具备多模数据处理与复杂查询能力。已广泛应用于金融、运营商、零售、互联网及公共服务等行业，助力 2000+ 客户实现关键业务系统升级。

### 助力科研及数据库人才培养



### OceanBase 数据库实验室

OceanBase 数据库实验室是探索科技前沿核心技术的发动机，与众多高校达成学术科研合作，还联合蚂蚁技术研究院、CCF 共同成立蚂蚁科研基金，重点在“单机分布式一体化”、“HTAP”、“存算分离”、“性能优化”、“AI 与 DB 融合”等方面进行进一步科学研究及成果转化。

**招聘岗位：**助理研究员（数据库方向）

**职位描述：**

- 探索业界领先的云原生分布式数据库技术，发表高水平论文或专利，提升 OceanBase 在数据库领域的业界影响力
- 将研究成果落地应用，解决 OceanBase 及外部客户的应用数据库的实际问题，满足业务需求
- 与国内外数据库领域的一流科研机构进行交流与合作

**技术方向：**

基于新型硬件的数据库技术研究、安全加密数据库、数据库隐私保护、高效分布式事务、事务与分析一体化、数据库智能与自治数据库。

**职位要求：**

- 在高水平国际会议或者学术期刊发表过相关论文，有数据库领域研究经验
- 良好的研究背景和成果，对数据库研究兴趣浓厚、业务抽象能力强，具备创新性思维
- 良好的团队协作及沟通能力，有一定编程能力，能够把研究想法转化为 Demo 应用
- 计算机科学、分布式系统、数据库等相关领域硕士、博士

**联系人：**

徐泉清（光启） E-mail : xuquanqing.xqq@oceanbase.com

OceanBase官网: [oceanbase.com](http://oceanbase.com)  
OceanBase开源社区: [open.oceanbase.com](http://open.oceanbase.com)



OceanBase 是100%根自研的原生分布式数据库，基于单机分布式一体化架构，高度兼容 Oracle 和 MySQL，支持事务处理 (TP)、实时分析处理 (AP) 和 AI 多工作负载融合，具备多模数据处理与复杂查询能力。已广泛应用于金融、运营商、零售、互联网及公共服务等行业，助力 2000+ 客户实现关键业务系统升级。

#### 根自研



“根自研”为关键业务系统兜底

#### 引领行业标准

##### 分布式开拓者

2010 年开始研发，作为全球最早投入分布式数据库研发的团队之一，唯一同时刷新 TPC-C 和 TPC-H 世界记录，历经超过 10 年金融与双十一等互联网大规模严苛场景的打磨与验证。

##### 权威认可

位居 IDC 中国分布式关系型数据库“领导者”象限，金融行业独立数据库厂商市场份额第一，产品能力国内第一，荣获 Gartner 云数据库亚太地区“客户之选”，并被评为全球“卓越表现者”。

##### 技术先进

业内首个单机分布式一体化架构，通过一个系统满足全生命周期数据存储与管理需求。凭借一体化产品设计，通过一个数据库支持 TP、AP、AI 等多种业务场景应用。

#### 产品体系

##### 1.企业版

适合于追求更高扩展能力、稳定性的中大企业

OceanBase 企业版 (OceanBase Database) 是一款完全自研的原生分布式数据库，在普通硬件上实现金融级高可用，首创“三地五中心”城市级故障自动无损容灾新标准，刷新 TPC-C 标准测试，单集群规模超过 1500 节点，具有云原生、强一致性、高度兼容 Oracle/MySQL 等特性。

##### 2.公有云

支持阿里云、华为云、腾讯云、AWS、Google Cloud 等多基础设施

OceanBase 公有云 (OceanBase Cloud) 是构建在阿里云、AWS 等全球主流公有云基础设施上，基于完全自主研发的原生分布式数据库，提供弹性扩展、卓越性能、主流兼容、高性价比的数据库云服务。为客户在云上提供服务、弹性、监控、诊断、开发、迁移、备份、恢复的端到端数据库服务化解决方案。

##### 3.社区版

适合于中小企业开发者

企业级开源分布式 HTAP (Hybrid Transaction/Analytical Processing) 数据库，具有原生分布式架构，支持金融级高可用、透明水平扩展、分布式事务、多租户和语法兼容等企业级特性。OceanBase 内核通过大规模商用场景的考验，已服务众多行业客户；未来将携手社区生态伙伴，共建开源开放的数据库内核和生态。

OceanBase官网: [oceanbase.com](http://oceanbase.com)  
OceanBase开源社区: [open.oceanbase.com](http://open.oceanbase.com)

## 25.5 长期铂金合作单位: 腾讯云数据库

 **腾讯云数据库**

# 腾讯云数据库TDSQL 企业级智能数据管理平台

◆ 产品矩阵及市场规模 ◆

TDSQL是腾讯云自主研发的企业级分布式数据库，其集中式和分布式两种类型数据库全部通过了中国信息安全管理中心安全可靠评测，在分布式计算、存储、元数据及管控方面实现了100%代码自研，是数据库国产化替代的最佳选择。

|           |                          |                              |                                  |                   |                   |
|-----------|--------------------------|------------------------------|----------------------------------|-------------------|-------------------|
| 应用于50万+客户 | 服务中国TOP10银行比例70%         | 整体收入及本地部署收入同比增速双料第一          |                                  |                   |                   |
| SaaS产品    |                          |                              |                                  |                   |                   |
| 关系型数据库    |                          |                              |                                  |                   |                   |
| 引擎产品      | 分布式数据库<br>TDSQL<br>MySQL | 云原生数据库<br>TDSQL-C<br>MariaDB | HTAP数据库<br>TDSQL-H<br>SQL Server | 数据库智能管家Dbbrain    | NoSQL数据库          |
|           | 云数据库<br>Redis            | 云数据库<br>MongoDB              | 云数据库<br>PostgreSQL               | 向量数据库<br>VectorDB | 云数据库<br>Memcached |
| 软硬一体化     | 云数据库独享集群                 |                              |                                  | 数据库一体机TData       |                   |

/ 登顶“双榜”世界第一，更高性能，更低成本，更安全稳定 /

|                 |                |               |                     |              |             |
|-----------------|----------------|---------------|---------------------|--------------|-------------|
| 刷新TPC-C记录       |                |               | 刷新TPC-DS记录          |              |             |
| 8.14亿<br>笔交易/分钟 | 99.999%<br>高可用 | 18秒<br>完成容灾切换 | 7260万QphDS<br>吞吐量第一 | 2倍<br>多人计算领先 | 1/3<br>价格减少 |

(\*TPC-C、TPC-DS是由国际事务处理性能委员会提供的数据库性能测评标准，以上记录分别在2023年、2024年打破)

◆ TDSQL：企业级智能数据管理平台 ◆

/ 赋能企业释放数据生产力 /

|            |           |
|------------|-----------|
| 金融级高可靠、高可用 | 领先的架构与性能  |
| 智能化运维管理    | 企业级数据安全合规 |

业务平滑迁移，国产化替代最佳选择

|    |                           |   |                                       |
|----|---------------------------|---|---------------------------------------|
| 政务 | 数字广东<br>Digital Guangdong | 四川省医疗保障局<br>Sichuan Medical Security Bureau | 河北人社厅<br>Hebei Social Security Bureau |
| 央企 | 中国邮政<br>CHINA POST        | 中国南方电网<br>China Southern Power Grid         | 中国中车<br>CRRC                          |
| 金融 | 中国银行<br>BANK OF CHINA     | 中国农业银行<br>Agricultural Bank of China        | 中国建设银行<br>China Construction Bank     |

DBbridge TDSQL 迁移工具箱(评估、迁移、验证)

内核层兼容 → 驱动层兼容 → 工具层兼容

/ 核心数据库替换最佳实践 /

|                     |                             |
|---------------------|-----------------------------|
| 平安银行                |                             |
| 高安全数据汇聚平台           | 平安银行信用卡系统国产化                |
| 人口库<br>110亿         | TDSQL规模达到<br>120+           |
| 法人主体<br>62亿         | 单个网关日间联机交易峰值<br>1.2万 笔请求/分钟 |
| 社会信用数据<br>61亿条      | 总数据量<br>100TB+              |
| 电子证照<br>累计签发 4.93亿张 | 夜间跑批峰值<br>98万 笔请求/分钟        |
| 地理要素<br>8亿          |                             |



# 助力打造国产数据库 产学研融合发展生态

◆ 产学研用深度合作，健全国产数据库研究生态 ◆

基于人大 - 腾讯联合实验室，产学研用联合创新

Tencent 腾讯 中国人大大学 中国科技大学 腾讯云 招商银行 平安银行 WeBank

产学研用协同创新 金融场景方案探索 科技成果产业化 产业生态推广建设

产学研用融合，助力高校人才培养

北京交通大学 中国人大大学 山东大学 南京大学 华中科技大学 中国科学技术大学 浙江大学

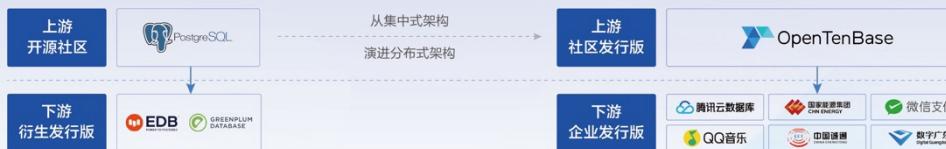
## 教育部协同育人计划、 教学内容和课程体系改革项目、师资培训项目

社区卓越贡献者：建设开放的、开发者友好的开源生态

TDSQLPG的数据库内核TBase在2019年11月对外开源，社区版本为OpenTenBase，已捐赠开放原子开源基金会。OpenTenBase开源代码仓库地址：Github.com/OpenTenBase。



10+ 国内外合作社区广泛案例应用：共建OpenTenBase社区发行版的下游生态。



## 25.6 长期铂金合作单位: 金篆信科

金篆信科 | GoldenDB®

金融、运营商核心系统市场排名第  
首批通过国测的分布式数据库 1

公司&产品简介

金篆信科有限责任公司是中兴通讯股份有限公司为拓展金融级交易型分布式数据库GoldenDB而成立的控股子公司，是中兴通讯着力培育和打造的新的重要增长极。

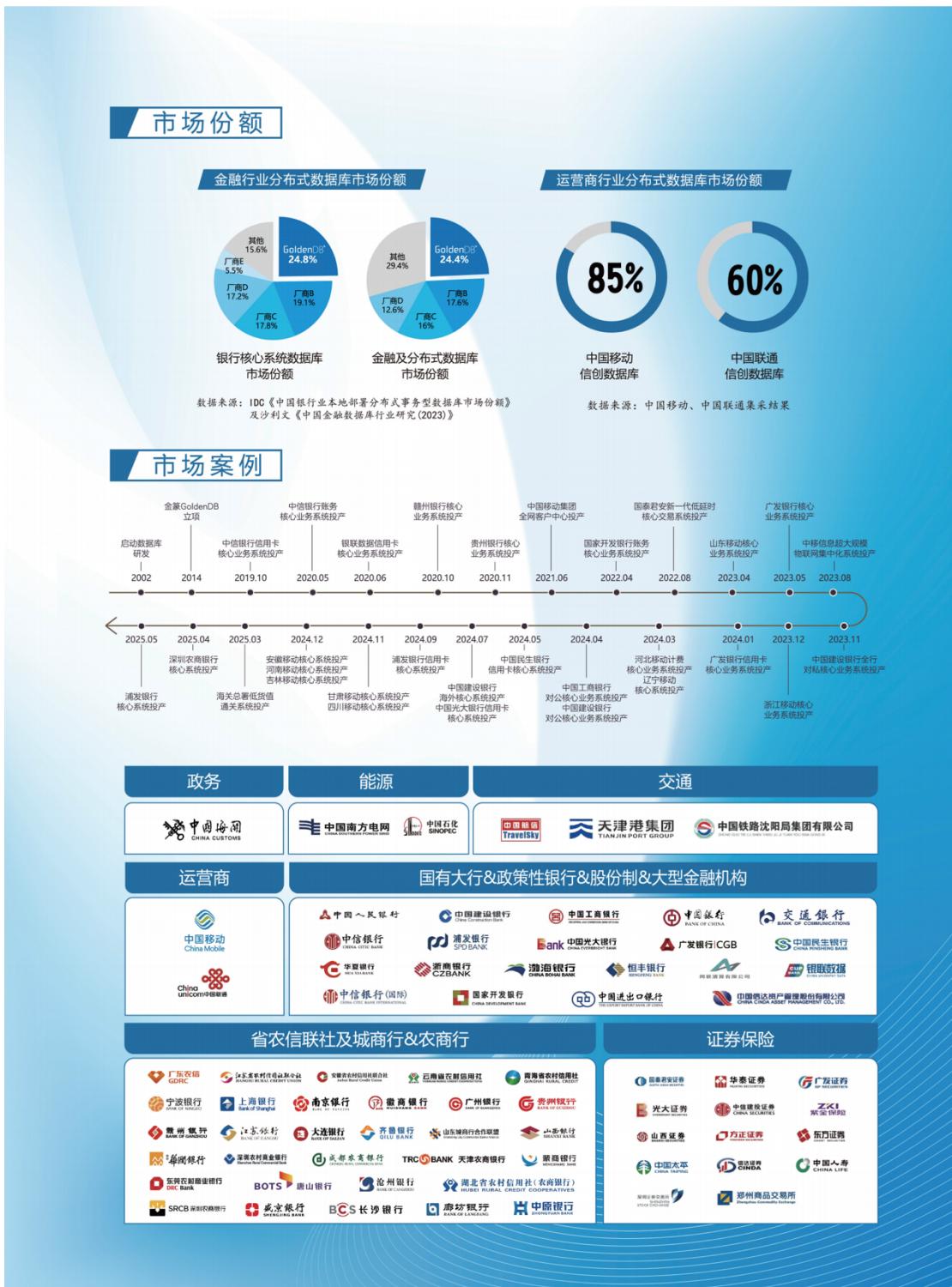
20+年  
研发积累      10+年  
重点行业锤炼      6.257亿+  
注册资金最大的数据库企业

700+员工  
研发人员占90%      800+  
数据库专利数量      双第一  
金融、运营商市场领先

每 日 承 载

|                  |                 |               |               |
|------------------|-----------------|---------------|---------------|
| 100亿+笔<br>金融核心交易 | 10万亿+元<br>金融交易额 | 9亿+<br>移动用户计费 | 12亿+<br>物联网计费 |
|------------------|-----------------|---------------|---------------|

GoldenDB®公众号



## 25.7 长期铂金合作单位: 阿里云





## CCF-阿里云瑶池科研基金

### ● 基金介绍

“CCF-阿里云瑶池科研基金”由阿里云计算有限公司与中国计算机学会（CCF）联合设立，专注于数据库领域，旨在为领域学者提供科研平台和必要的资源支持，鼓励优秀学者揭榜挂帅，开展与产业相结合的前沿探索性工作，解决真实世界中的卡脖子难题。首批8个项目在2024年正式启动。

### ● 课题方向

2025年，基金将围绕Data+AI时代数据处理与分析的新问题设立10项研究课题，欢迎海内外高校科研团队申请！

**AI时代的数据库系统：多模处理 + 实时分析，像乐高一样易用、好用**

10项课题 | 30万经费支持 | 最高40万元阿里云计算资源优惠



### ● 联系方式

基金咨询邮箱: [apsaradb\\_fund@alibabacloud.com](mailto:apsaradb_fund@alibabacloud.com) 敬请期待

## 2025全国大学生计算机系统能力大赛 ——第2届PolarDB数据库创新设计赛

### ● 赛事介绍

2025全国大学生计算机系统能力大赛——第2届PolarDB数据库创新设计赛是由系统能力培养研究专家组发起、由全国高校计算机教育研究会和系统能力培养研究项目发起高校主办、阿里云承办，面向高校大学生的全国性大赛。比赛以赛促学、以赛促教，鼓励学生设计、实现综合性的数据库系统，提升学生技术创新、工程实践、团队协作能力，助力我国高校与企业在数据库系统领域的产学研合作健康、快速发展。



丰厚奖励



大赛证书



面试速通



工程技术实战

引领创新，挑战极限！无论你是数据库新手，还是技术大咖

2025年全国大学生计算机系统能力大赛——第2届PolarDB数据库创新设计赛邀你来参加！

### ● 赛题简介

向量计算已成为数据库领域中最常见的计算任务之一。PolarDB PostgreSQL版（简称PolarDB PG）是一款阿里云自研的企业级数据库产品，100%兼容PostgreSQL。PolarDB PG结合PGVector插件，通过自定义的数据类型和存储方法，使高维向量计算变得更加高效、快速。比赛希望参赛者通过创新的优化方法，充分挖掘数据库系统AI潜力，推动数据库向智能化时代跃迁。



比赛报名 & 赛题相关信息将于近期公布，敬请期待！

欢迎微信扫码关注「阿里云瑶池数据库」公众号

第一时间获取大赛动态信息

## 25.8 长期铂金合作单位: TDengine

### TDengine IDMP AI原生的工业数据管理平台

基于采集的数据, TDengine 自动感知应用场景, 无需提问, 就自动生成看板、报表, 自动进行实时数据分析, 自动监测异常, 让您拥有一个工业智能体, 实现零依赖的实时决策革命。



#### 为什么要选择 TDengine IDMP



##### 数据自己说话

无需您提问, 基于采集的数据, TDengine IDMP 利用 LLM, 自动感知应用场景, 自动生成场景特有的指标、可视化看板、报表和实时数据分析。无需业务知识的多年积累, 无需主动查询, 核心洞察主动推送。



##### 决策零等待

除 AI 主动推送的看板、分析之外, 您还可以用自然语言主动提问与数据相关的任何问题。无需数据分析师、IT 工程师的帮助, AI 基于采集的数据实时给予答案, 您即可形成行动方案。从提问到决策, 分钟级闭环。



##### 全栈解决方案

与 TDengine 时序数据库一起, 为工业数据管理, 提供从数据采集、清洗、情境化、标准化, 到存储、查询、实时分析、预测、异常检测, 再到可视化、事件管理等全栈的解决方案。架构极简, 运维轻量化。



##### 开放的企业级应用

提供单点登录, 提供基于角色的权限控制、提供数据模型的版本控制, 支持数据备份、异地容灾和实时分发, 支持虚机、容器部署, 支持 Windows 与 Linux, 能与 MES、ERP、AI 等企业应用系统无缝集成。

#### 零风险、零初期成本, 立即体验

完全零代码, 零 SQL, 无学习成本。无论是本地部署还是云服务, 都可以免费体验全部功能和性能, 验证自己的场景。

系统自带的示例场景, 更是能帮您迅速验证自己的想法, 完全零门槛。



免费下载体验



免费注册云服务

## 应用场景

### 物联网

智能硬件、大型联网装备、智慧家庭、智慧城市、智慧矿山、智慧水务、智慧农业

### 车联网

TBox 数据处理、新能源汽车电池数据、自动驾驶、智慧物流、车队管理

### 工业互联网

智能制造、流程行业、生产数字化、数字孪生

### 电 力

风力发电、光伏发电、配电、用电、能源管理

### IT运维

IDC、主机、虚机、网络设备、容器、微服务监测

### 金 融

股票、证券、期货行情数据

## 合作客户



「合作伙伴与客户」



## 六大功能

### 数据建模

Industrial Data Modeling

### 数据情景化

Data Contextualization

### 数据标准化

Data Standardization

### 智能的数据可视化看板

Industrial Data Visualization

### 智能的实时数据分析

Real-time Analytics

### 事件管理与根因分析

Event Management

www.taodata.com

4006120020

13121517901 (企业微信)



官方网站

微信公众号

企业微信

视频号

## 25.9 长期铂金合作单位: 金仓数据库

**KING BASE** | 数据库平台用金仓

学科 自主开创 | 产品 自主研发 | 人才 自主培养

公司简介

中电科金仓(北京)科技股份有限公司(原人大金仓)成立于1999年,是我国成立最早的拥有自主知识产权的数据库企业,也是中国电子科技集团(简称“CETC”)成员企业。金仓数据库自成立以来坚持自主创新,专注数据库领域20余载,具备出色的数据库产品研发及服务能力,核心产品金仓数据库管理系统KingbaseES(简称“KES”)是具备国际先进水平的大型通用融合数据库。

金仓数据库广泛服务于能源、金融、电信、交通、医疗、政务等多个重点行业和关键领域,累计装机部署超过100万套。自2020年起,金仓已连续四年在国产数据库关键应用领域销售套数占比第一;2022年,金仓数据库在中国石油石化行业和交通行业销售量居国内厂商第一位置;2023年,金仓数据库在中国医疗行业和交通行业销售量居国内厂商第一位置。

金仓数据库在北京、成都、天津、青岛、西安等地设有研发和服务中心,服务网络覆盖31个省、自治区、直辖市及港澳地区,保证了7×24小时优质原厂的本地化服务能力,为客户提供全面的服务和安全保障。未来,金仓将持续践行数据库领域国家队使命,夯实基础软件底座,踔厉奋发,笃行不怠,为各行业数字化场景提供数据的存储与计算支撑,推动新质生产力发展,为我国数字经济建设筑牢坚实基石。

\* 数据来源: 赛迪顾问2020-2024年《中国平台软件市场研究年度报告》

### “核”计划 —— 打造国产数据库开发者的硬核能力 >>>

金仓数据库基于KES数据库产品,围绕数据存储与存取,查询处理与优化、事务处理与并发控制等数据库管理系统的相关内容,学习和掌握关系型数据库管理系统的实现技术(包括基本原理、算法和实现技术等)。通过理论学习和模块开发训练帮助学员用系统、全局的观点来看待复杂软件系统实现;助力学员理解分而治之、分层实现在复杂软件系统中的重要作用,为数据库内核研发领域培养更多的人才。

#### 部分合作高校 >



中电科金仓(北京)科技股份有限公司

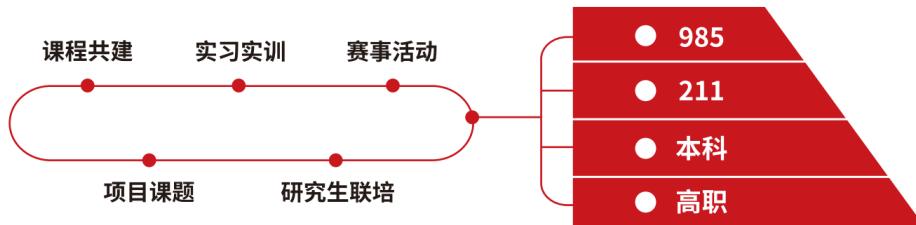
全国服务热线: 400-601-1188

北京市朝阳区容达路7号E座

联系人: 13683386196 (巩经理)



## ➤ 校企合作方案



| 类别    | 共建方式  | 课件                      | 工具           | 服务   |
|-------|-------|-------------------------|--------------|------|
| 数据库实现 | 专业核心课 | 《数据库系统原理与实现》            | KES+RUCBASE  | 安装部署 |
|       |       | 金仓数据库内核开发课程             | RUCBASE      | 线下授课 |
| 数据库应用 | 实验课   | 《数据库系统概论实验指导书》          | KES          | 师资培训 |
|       | 应用开发课 | 《金仓数据库快速应用开发》           | KES+KES PLUS | 师资培训 |
| 数据库运维 | 专业选修课 | 《金仓数据库KingbaseES DBA实战》 | KES          | 师资培训 |
|       | 实训认证课 | 金仓数据库DBA培训认证            | KES          | 认证考试 |
| 思政讲座课 |       | 《数据库产业发展趋势》             | —            | 现场讲座 |

## ➤ 实验室建设方案

- 01 | 课程资源**
  - 数据库系统概论实验课程
  - 金仓数据库KingbaseES DBA实战课程
- 02 | 软件产品**
  - 金仓数据库管理系统(KingbaseES)
  - 金仓快速开发与运维平台(KES Plus)
  - 金仓云数据库一体机(KXData-M)
- 03 | 培训服务**
  - 数据库管理工程师 • 初中级培训 • 高级培训
  - 国产数据库方向的赛前培训辅导
- 04 | 其他服务**
  - 课程思政



中电科金仓(北京)科技股份有限公司

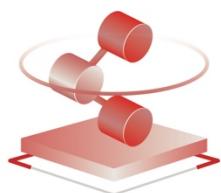
北京市朝阳区容达路7号E座

全国服务热线: 400-601-1188

## 25.10 长期铂金合作单位：南大通用



### | 核心产品



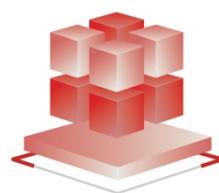
**GBase 8a**

分布式逻辑数据仓库



**GBase 8s**

基于共享存储的数据库集群



**GBase 8c**

多模多态数据库管理系统

### | 他们都在使用GBASE数据库

|                                      |                                      |                                   |  |                              |
|--------------------------------------|--------------------------------------|-----------------------------------|--|------------------------------|
| 财政部                                  | 人社部                                  | 工信部                               | 发改委                                      | 国家税务总局                       |
| 中国人民银行<br>THE PEOPLE'S BANK OF CHINA | 中国农业银行<br>AGRICULTURAL BANK OF CHINA | 中国建设银行<br>China Construction Bank | 中国邮政储蓄银行<br>POSTAL SAVINGS BANK OF CHINA | 招商银行<br>CHINA MERCHANTS BANK |
| PICC<br>中国人民保险                       | 阳光保险集团<br>Sunshine Insurance Group   | 平安银行<br>PING AN BANK              | 广发银行                                     | 渤海银行<br>CHINA BOHAI BANK     |
| 中国移动<br>China Mobile                 | 中国联通<br>China unicom                 | 中国电信<br>CHINA TELECOM             | 国家电网                                     | AFC                          |
| 中国石油                                 | 中国石化<br>SINOPEC                      | 国家电网<br>STATE GRID                | 国家管网<br>PipeChina                        | 中国烟草<br>CHINA TOBACCO        |



专注于数据库软件产品和服务

### | 社招高端人才

#### 高端研发工程师 (C/C++)

📍 天津、南京、西安、济南

#### 行业销售总监

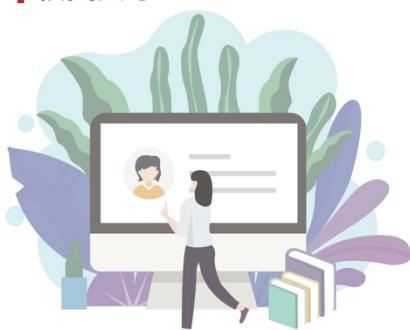
(金融、政企、运营商、交通、能源、电力、发电等行业)

📍 北京、天津、上海、南京、杭州、广州、  
深圳、成都、银川、西安、福州、海口



✉ 社招简历投递: gbasehrrc@gbase.cn  
(主题: 姓名+岗位+base地)

### | 校园人才



#### 研发工程师 (C/C++)、测试工程师、 产品工程师、技术服务工程师

📍 天津、南京、西安、济南

##### 校园招聘流程



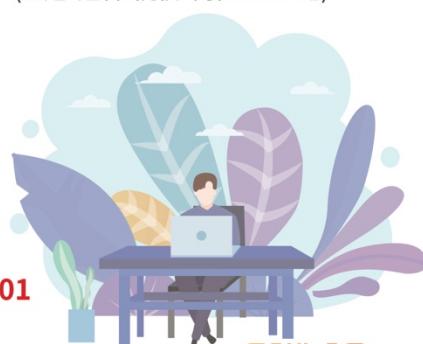
✉ 校招简历投递: gbasehrxy@gbase.cn  
(主题: 姓名+院校+岗位+base地)

### | 博士后工作站

#### 研究方向

- › 云原生环境下的高性能分布式数据库研究
- › 软硬件一体化高性能的数据仓库研究
- › 数据库及数据仓库关键算法研究

博士后进站咨询: 王先生022-58815678-8001



南大通用数据技术股份有限公司 General Data Technology Co.,Ltd.

商务热线:  
400-013-9696

官方网址:  
[www.gbase.cn](http://www.gbase.cn)

公司邮箱:  
[info@gbase.cn](mailto:info@gbase.cn)



GBASE数据库公众号

## 25.11 长期铂金合作单位: 移动云

中国移动 | 移动云 大云海山®

# 大云海山数据库

移动云数据库紧跟中国移动云能力中心转型、升级,历经十余年技术迭代淬炼,在数据领域积累深厚底蕴,专注云数据库产品研发。目前已构建完整产品和市场体系,涵盖集中式、分布式、数仓、NoSQL等各类数据库管理系统以及管理平台。研发团队和技术能力经历了大规模商用化的检验,持续坚定自主研发,成功孵化大云海山数据库技术品牌,内核技术自主掌控,同时积极生态技术创新。

### 移动云一站式数据库产品体系

| 党政   | 教育  | 电信  | 金融 | 医疗 | 航天航空   | 石油 | 交通 | 电力 | 泛互联网 |
|--|-----|-----|----|----|--|----|----|----|------|
| 公有云  | 私有云 | 边缘云 |    |    |  |    |    |    |      |
| 大云海山 云原生   |     |     |    |    | 云托管  |    |    |    |      |
| 云原生数据库 PostgreSQL版<br>云原生数据库 分布式版<br>云原生数据库 数仓版<br>多模数据库*<br>(图、时序等) |     |     |    |    | RDS数据库<br>(MySQL、PostgreSQL、达梦、TiDB、SQL Server、金仓)<br>云数据库Redis<br>文档数据库服务<br>搜索数据库服务<br>向量数据库<br>(ES、MySQL、PG、Redis等) |    |    |    |      |
| 云腾海山 数据库生态工具   |     |     |    |    | AI   |    |    |    |      |
| 数据库传输<br>数据库管理   |     |     |    |    |  |    |    |    |      |
| 数据库统一云原生平台底座   |     |     |    |    |  |    |    |    |      |
| 场景   |     |     |    |    |  |    |    |    |      |
| 中小企业上云<br>国产化改造<br>人工智能<br>数据安全<br>行业转型                              |     |     |    |    |  |    |    |    |      |

已注册大云海山®、云行海山®、云腾海山®

### 客户案例

- 央教馆**  
国家教育数字化平台全面上云  
支撑40W+人  
在线观看《开学第一课》  
100MB/s  
40T数据快速上云
- 139邮箱**  
Oracle数据库国产化替换  
9亿+用户  
80TB数据量  
95%以上  
Oracle兼容性
- 移动云盘**  
智能数据检索加速  
“AI+”升级  
降低30%↓  
运维成本  
秒级响应  
大幅缩短检索时间
- 吉威空间**  
自然资源与地理空间基础信息库  
百TB数据  
海量存储  
QPS提升44%↑  
增强业务负载能力

## 产品服务与应用场景

### 高阶服务

覆盖从初创企业到国家级平台各类企业全生命周期数据库需求, 提供一站式数据库服务

SLA 99.95%, 业务连续性保障

兼容业界主流技术栈, 适配工作量少

服务、备份、运维等一体化服务, 降低用户TCO

一键迁移上云, 数据库秒级割接



核心需求  
高可用性、强一致性、全球化部署

解决方案

大容量、分布式、智能化

高阶能力供给  
高可靠



核心需求  
成本控制、弹性扩展、高频迭代

解决方案

Serverless弹性架构、镜像热升级

弹性敏捷  
高利用率



核心需求  
按需付费、极简运维、快速上线

解决方案

自动化运维、专业工作台

一站式服务  
低成本

### 信创资质

面向党政国企、金融医疗等行业, 提供全生态、全链路、全场景信创数据库解决方案

全面适配国产化基础设施底座

产品互认证52项, 覆盖信创“6+3”

芯片



操作系统



信创安全 行业认证

等保三级(S3A3)防护, 多产品通过可信云认证、  
江苏省首版次、中央政府采购名录、信创图谱

安全体系

▶ 4大原生能力

用户认证、身份管理、访问控制、数据加密

▶ 10大增强能力

SM4国密算法、透明加密、数据脱敏、差分隐私保护、...

### 智算能力

向量数据库助力企业快速构建智能检索、智能推荐、大模型训练等AI场景, 释放非结构化数据价值



海量+多模数据的实时检索

通过向量与标量的混合检索与分页查询策略, 成功实现对时间、地点、任务、标签及图片内容多模态数据联合检索, 支持千亿级数据毫秒级响应。

大模型+RAG场景的核心引擎

相比传统场景, 可有效避免大模型幻觉, 避免高昂模型微调成本。问题回答更精准、更高效、更安全。

「首个」

业界向量数据库标准发布

「首批」

通过中国信通院测评

「国家级」

入选工信部典型案例

## 25.12 长期金牌合作单位: 海量数据


**海量数据**  
 VASTDATA | **专注做好数据库**

北京海量数据技术股份有限公司（股票代码：603138.SH）成立于2007年，是国内首家以数据库为主营业务的主板上市企业。公司十余年来秉承“专注做好数据库”的初心，始终致力于数据库产品的研发、销售和服务。核心产品海量数据库Vastbase系列、数据库一体机Vastcube系列、存储Vaststorage系列，全栈国产化，应用满足度高，目前广泛应用于政务、制造、金融、通信、能源、交通等多个重点行业，已成为国产企业级数据库的首选之一。

**6大**  
研发中心

**65%**  
技术精英

**7大**  
服务中心

**30+**  
城市覆盖数

**产品体系**
**部分案例**

**海量数据库 Vastbase**

|       |         |        |
|-------|---------|--------|
| 数据库产品 | 数据库生态工具 |        |
| G100  | VDS     | exBase |
| V100  | VEM     | vCloud |

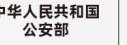
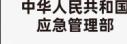
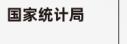
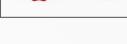
**海量计算 Vastcube**

|                        |                        |                        |
|------------------------|------------------------|------------------------|
| 数据库一体机<br>Vastcube G系列 | 数据库一体机<br>Vastcube E系列 | 数据库一体机<br>Vastcube V系列 |
| G1000                  | E1000                  | V3000                  |
| G3000                  | E3000                  |                        |

**海量存储 Vaststorage**

|                  |                  |                 |
|------------------|------------------|-----------------|
| 混合闪存存储<br>1000系列 | 全闪存存储<br>1000H系列 | 分布式存储<br>2000系列 |
|------------------|------------------|-----------------|

**部分案例**

|   |   |   |
|---|---|---|
|    |    |    |
|   |   |   |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

北京市海淀区学院路30号科大天工大厦B座6层 | 400-012-3005 | [www.vastdata.com.cn](http://www.vastdata.com.cn)



## 25.13 长期金牌合作单位: 百度

百度智能云数据库致力于为企业用户打造稳定、高效的一站式数据库产品和服务，包括关系型/非关系型数据库、数据仓库，数据库生态工具等，满足企业业务全生命周期的数据生产、管理、运维需求。



百度智能云数据库产品矩阵

### 云原生数据库 GaiaDB

- 100%兼容MySQL，平滑迁移零成本
- 跨可用区容灾，RTO小于30秒，服务可用性达99.99%
- 复杂查询性能超MySQL百倍
- 总拥有成本降低50%，高性价比企业首选

### 向量数据库 VectorDB

- 自研引擎架构，支持百亿向量存储
- 查询性能领先，QPS高出开源方案7倍以上
- 支持LangChain、MemoO、Dify、FastGPT等主流工具链，兼容上下游生态，集成灵活无缝
- 荣获 2024年IDC向量数据库报告Top1

扫描下方二维码秒速关注产品信息和动态，新用户即刻领取免费实例，还有0.1折首购优惠活动哈！



VectorDB



GaiaDB

## 25.14 长期金牌合作单位: 字节跳动



### ByteDance Technical Infrastructure Team

The ByteDance Technical Infrastructure team is mainly responsible for the company's cloud infrastructure construction and supports multiple APP products such as Douyin, Toutiao, Tomato Novel, Xigua Video, CapCut, etc. And it also supports the public cloud business of Volcano Engine. To date, we manage hyper-scale data centers through cloud technology. We manage more than ten million container instances through our deeply optimized Kubernetes to support 100,000+ micro-services. We also govern EB-level data assets through our storage products matrix such as NewSQL, NoSQL, and cloud storage. At the same time, we actively embrace open source and innovative software & hardware architectures, and are committed to building industry-leading cloud infrastructure.

#### Infrastructure System Lab

The Infrastructure System Lab in the Technical Infrastructure team works on cutting-edge infrastructure system innovations with a focus on intersection of AI/LLM with infrastructure systems. Along with the development teams, the lab leads state-of-the-art technical breakthroughs in fields that then lead to the creation of powerful backbones for ByteDance products and corporate partners.

Our lab actively works on research in the following major technical areas:

- AIOPS, AI for OR, AI/LLM for Infrastructure (DB, Storage, Networking...)
- Cutting-edge data management systems and technologies, such as Vector DB, Multi-modal AI native DB, HTAP, Trusted DB, Autonomous DB, TextToSQL, etc;
- High-performance and resource-efficient key value stores
- Distributed KV cache for LLM serving acceleration

Besides significant business impact, our lab also encourages publishing in top-tier conferences such as SIGMOD, VLDB, FSE, ICLR, EuroSys etc and filing patents. In 2025, we already have 10+ papers accepted by top tier conferences.

#### Our Positions

Experienced



Campus Hiring

智能系统研究员 - 基础架构  
- 筋斗云人才计划



© 北京

© 杭州

## 25.15 长期金牌合作单位: 崖山数据

深圳计算科学研究院  
Institute of Computing Sciences
 崖山数据库  
YashanDB

### 崖山数据库YashanDB

首席科学家樊文飞院士领军打造  
实力雄厚的研究团队+产业经验丰富的工程团队

|      |       |        |
|------|-------|--------|
| 500+ | 121篇  | 80%    |
| 员工人数 | 高水平论文 | 研发人员占比 |

深圳计算科学研究院是深圳市人民政府批准建设的“十大基础研究机构”之一，以建设世界一流的计算基础科学研究院为目标，致力于打破国外技术垄断，重点研究和开发以大数据为代表的基础理论、底层技术、核心系统，打造中国原创、具有源头技术的基础软件品牌，引领计算科学产业创新发展。

崖山数据库系统YashanDB是深圳计算科学研究院重点自主设计研发的新型数据库系统，提供包含读写分离集群、共享集群、空间数据库、分布式实时数仓等系列数据库产品及解决方案，覆盖OLTP/HTAP/OLAP交易和分析混合负载场景，全面兼容私有化及云基础设施，为客户提供一站式的企业级融合数据管理解决方案。

**科研为立身之本 做中国自己的数据库品牌**

**原创资源受限计算理论**  
打破大规模数据分析算力边界，把“大数据变小”，实现最高5个数量级的性能提升。

**Oracle高端场景“1:1”平替**  
攻克数据库技术“制高点”共享集群技术，实现在高端核心业务系统对国外数据库的“1:1”平替。

**推动产学研用融合 协同筑牢高校人才根基**

**携手前沿高校，深度参与课程共建**  
• 与深圳大学，开发《数据库系统》课程  
• 与深圳大学，走进《数据库内核》课堂  
• 与天津大学，启动《数据库应用实践》课程联合开发计划

**3级认证培训体系，权威、严谨、实用**  
• 完成认证培训10期+  
• 生源遍布省份20+  
• 持证人才4100+

**携手10+高校，建立高质量实习就业合作机制**

**大力推动SQLab平台建设**  
• 提供开箱即用的YashanDB云环境  
• 支持SQL实时执行、笔记分享  
• 服务用户1700+

**典型用户**

关注微信公众号

高校合作沟通

## ■ 26. 承办单位介绍



吉林大学是教育部直属的全国重点综合性大学，坐落在吉林省长春市。学校始建于 1946 年，1960 年被列为国家重点大学，1984 年成为首批建立研究生院的 22 所大学之一，是国家“211 工程”“985 工程”和首批“双一流”大学建设高校。

学校始终秉持“求实创新、励志图强”的校训，坚持以“学术立校、人才强校、创新兴校、开放活校、文化荣校”为发展战略，努力建设成为在国家和区域经济社会发展中具有重要地位的高素质创新人才培养、高水平科学研究和成果转化、高质量社会服务、高水平国际交流合作、先进文化引领的重要基地；成为让学生全面发展、让教职工引以自豪、让社会高度赞誉、让世界广泛认同的大学。到建校 100 周年时，努力把吉林大学建成中国特色、世界一流大学。

学校学科门类齐全，下设 48 个教学单位，涵盖哲学、经济学、法学、教育学、文学、历史学、理学、工学、农学、医学、管理学、艺术学、交叉学科等 13 大学科门类；有本科招生专业 121 个，一级学科博士学位授权点 49 个，一级学科硕士学位授权点 6 个，博士专业学位授权点 11 个，硕士专业学位授权点 25 个，新兴交叉学科学位授权点 8 个；博士后科研流动站 48 个。19 个学科（领域）的 ESI 排名进入全球前 1%，其中 5 个学科排名进入全球前 1‰，1 个学科排名进入全球前 1‰。

学校师资力量雄厚，有专任教师 6283 人，其中教授 2444 人。中国科学院和中国工程院院士 12 人，哲学社会科学资深教授 7 人，国务院学位委员会学科评议组成员 21 人，吉林大学各类国家级人才总数为 529 人次，吉林省“长白山学者”和“长白英才计划”人才在岗 175 人，吉林大学“匡亚明/唐敖庆学者”在岗 998 人。

学校现有全国重点实验室 7 个，国家重大科技基础设施 1 个，国家大学科技园 1 个，“一带一路”联合实验室 1 个，国家工程实验室 1 个，国家地方联合工程实验室 6 个，国家工

程技术研究中心 1 个，国家国际科技合作基地（国际联合研究中心）4 个，国家应用数学中心 1 个，国家高端智库 1 个，国家 2011 协同创新中心（核心协同）2 个，铸牢中华民族共同体意识研究基地 1 个，国家革命文物协同研究中心 1 个，教育部人文社会科学重点研究基地 6 个，教育部哲学社会科学实验室 1 个，教育部重点实验室/研究中心 18 个，教育部国际合作联合实验室 2 个，教育部省部共建协同创新中心 4 个，教育部野外科学观测研究站 1 个，教育部医药基础研究创新中心 1 个，其他行业部委重点实验室 25 个，吉林省实验室 1 个。学校承担了大量国家级和省部级科研项目，产出了一批产业化前景好、技术含量高的高技术成果。

学校已建立起学士—硕士—博士完整的高水平人才培养体系。在籍学生 75413 人（全日制 72112 人、非全日制 3301 人），其中预科生 151 人，本科生 41939 人，硕士生 22805 人（全日制 19974 人、非全日制 2831 人），博士生 9661 人（全日制 9191 人、非全日制 470 人），留学生 853 人。

学校聚焦世界名校合作，逐步完善全球合作网络，目前已经与 43 个国家和地区的 313 所高校和科研机构建立了合作关系。其中，排名世界前 100 名的 50 所，前 200 名的 76 所。学校与 19 个国家的高校和科研机构合作共建了 49 个中外合作平台。

学校校园占地面积 734.58 万平方米（其中珠海研究院 150.35 万平方米），校舍建筑面积 292.61 万平方米。学校在珠海市建有珠海研究院，占地面积 150.35 万平方米。学校图书馆各类藏书 859.19 万册，是中国高等教育文献保障系统（CALIS）东北地区中心，大学数字图书馆国际合作计划（CADAL）项目成员馆，中国高校人文社会科学文献中心（CASHL）东北区域中心。

在建设世界一流大学的进程中，吉林大学将努力做到在关心国家命运、服务国家战略上有所作为，让党和国家满意；在勇担社会责任、满足社会对优质高等教育不断提高的要求上有所进步，让广大人民群众满意；在坚持以人为本、实现好维护好发展好学校广大师生员工根本利益上有所建树，让广大师生员工满意。



## 27. 会务信息

### 27.1 会议会场

会场：吉林长春国际影都万达文华酒店

会议官方网址：<https://ccf.org.cn/ndbc2025>

### 27.2 会议交通及住宿

#### 27.2.1 乘车指南

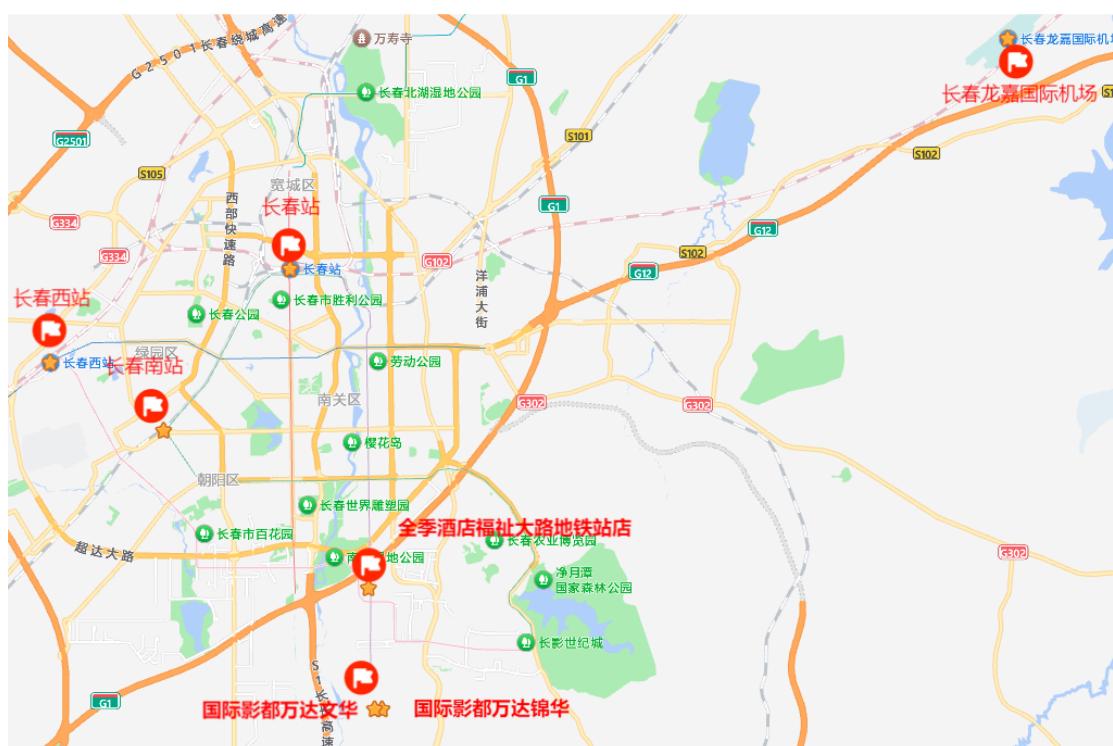


图 1 吉林长春国际影都万达文华酒店地址示意图

吉林长春国际影都万达文华酒店

地址：吉林省长春市南关区净月高新区文远大路 1555 号

电话：0431-89178888

### 一、长春龙嘉国际机场至长春国际影都万达文华酒店、锦华酒店：

- 1、乘出租车：预计行驶49公里，全程约45分钟，单程费用约85元。
- 2、公共交通：机场巴士2号线→地铁4号线(龙嘉国际机场站上车，往卫星广场方向，临河街站下车，步行300约300米，职业学院站A口换乘4号线，天新路站下车，B2口出，步行约1.2公里。 (全程约1小时54分钟)

### 二、长春南站至长春国际影都万达文华酒店、锦华酒店：

- 1、乘出租车：预计行驶21公里，全程约40分钟，单程费用约55元。
- 2、公共交通：地铁3号线→地铁4号线(南站步行至孟家屯站上车，职业学院站换乘4号线，天新路地铁站下车B2口出，步行约1.2公里， (全程约1小时14分钟)

### 三、长春站至长春国际影都万达文华酒店、锦华酒店：

- 1、乘出租车：预计行驶27公里，全程约40分钟，单程费用约68元。
- 2、公共交通：地铁4号线(长春站北站上车，天新路站下车B2口出，步行约1.2公里， (全程约1小时8分钟)

### 四、长春西站至长春国际影都万达文华酒店、锦华酒店：

- 1、乘出租车：预计行驶30公里，全程约50分钟，单程费用约79元。
- 2、公共交通：地铁6号线→地铁4号线(长春西站上车，长影世纪城方向，福祉大路站（站内）换乘4号线，天新路站下车，B2口出步行约1.2公里， (全程约1小时8分钟)

## 27.2.2 住宿环境

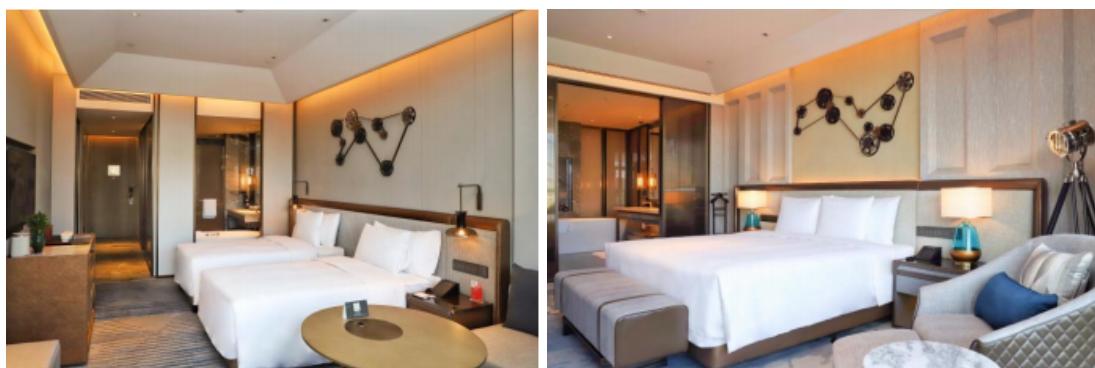
本次会议提供协议酒店共三家，住宿环境如下：

### (1) 长春国际影都万达文华酒店（五星级）



长春国际影都万达文华酒店于2024年5月25日正式启幕。作为长春国际影都项目的重要组成部分，酒店以“光影交错，齿轮间的艺术之旅”为设计主题，巧妙融合齿轮机械与装饰艺术派元素，打造出一个充满艺术气息和影视文化的空间。酒店不仅为电影相关人士提供高品质的住宿服务，也为游客带来独特的影视文化体验。酒店拥有167间典雅的豪华客房和奢华套房，采用了融合了电影元素的时尚设计装潢。客房拥有落地窗和步入式私家阳台，可欣赏到花园景观及城市景色。

| 住宿房型 | 优惠价           |
|------|---------------|
| 标准间  | 550元/间/天（含双早） |
| 大床房  | 550元/间/天（含单早） |



文华酒店标准间（左）、大床房（右）

(2) 长春国际影都万达锦华酒店（四星半）



万达锦华酒店拥有311间时尚舒适的客房和套房，采用了融入了电影元素的新潮装潢设计，服务设施功能齐全，透过落地窗可欣赏花园景观及城市景色。

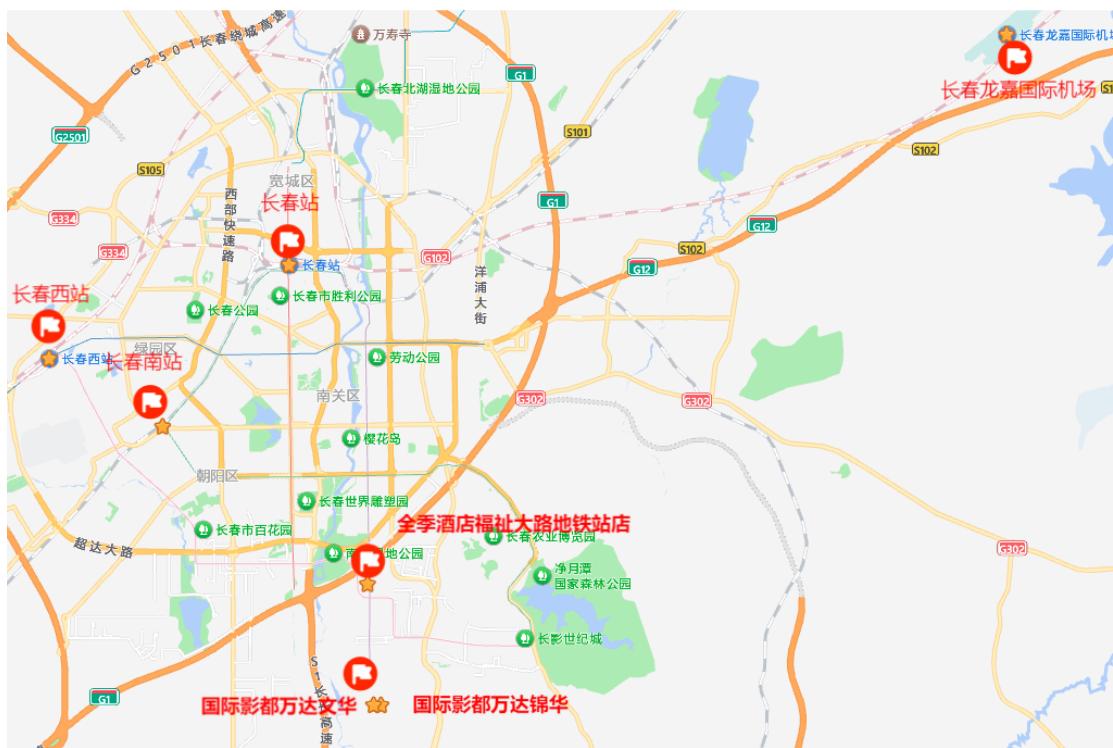
| 住宿房型 | 优惠价           |
|------|---------------|
| 标准间  | 450元/间/天（含双早） |
| 大床房  | 450元/间/天（含单早） |



锦华酒店标准间（左），大床房（右）

### (3) 全季酒店（长春福祉大路地铁站店）

地址示意图



地址：吉林省长春市南关区临河街10555号(距天工路地铁站200米)

电话：0431-81310555

该酒店距离会议酒店直线距离5.5公里，驾车14分钟，出租车单程费用约16元，公共交通乘坐地铁4号线，全程约42分钟

#### 一、长春龙嘉国际机场至全季酒店（长春福祉大路地铁站店）：

- 1、乘出租车：预计行驶40公里，全程约45分钟，单程费用约80元
- 2、公共交通：机场巴士2号线→地铁4号线(龙嘉国际机场站上车，临河街站下车，步行约300米，职业学院站A口换乘4号线，天工路站下车A口出，步行约710米。全程约1小时36分钟)

#### 二、长春南站至全季酒店（长春福祉大路地铁站店）：

- 1、乘出租车：预计行驶16公里，全程约30分钟，单程费用约37元
- 2、公共交通：地铁3号线→地铁4号线(南站步行450米至孟家屯站上车，职业学院站换乘4号线，天工路站下车A口出，步行约710米，全程约56分钟)

#### 三、长春站至全季酒店（长春福祉大路地铁站店）：

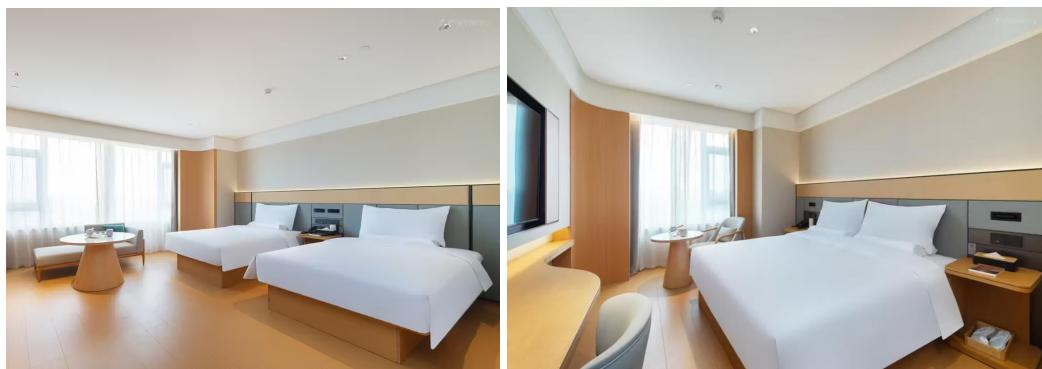
- 1、乘出租车：预计行驶20公里，全程约40分钟，单程费用约48元
- 2、公共交通：地铁4号线(长春站北上车，天工路站下车A口出，步行约710米，全程约50分钟)

#### 四、长春西站至全季酒店（长春福祉大路地铁站店）：

- 1、乘出租车：预计行驶25公里，全程约50分钟，单程费用约70元
- 2、公共交通：地铁6号线(长春西站上车，福祉大路站下车A口出，步行约1.1公里，全程约52分钟)



| 住宿房型 | 优惠价           |
|------|---------------|
| 标准间  | 380元/间/天（含双早） |
| 大床房  | 380元/间/天（含单早） |



全季酒店标准间（左），大床房（右）