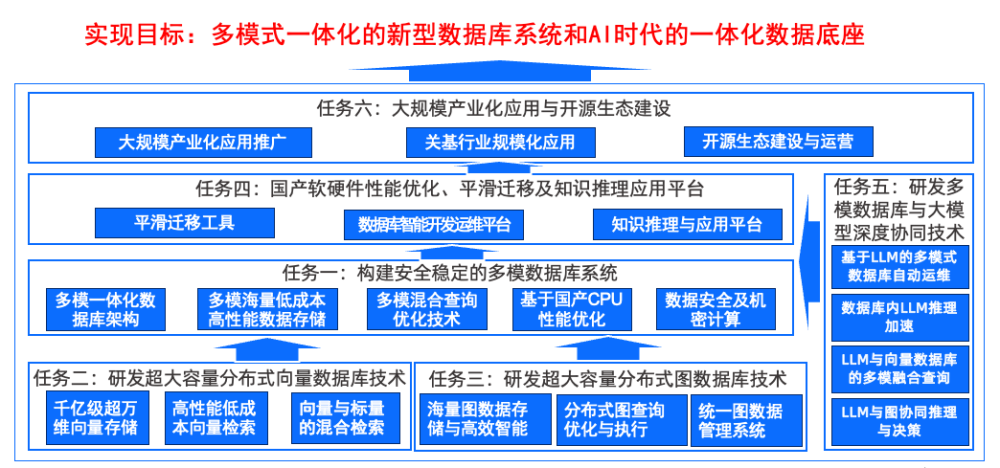
# 任务二



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 研究内容 | 指南要求 | 指标（待精简） | 已有数据（作为指标参考） | 修改后指标 |
| 子课题1: 海量图数据存储与高效智能索引 | 面向 PB 级的高性能低成本存储引擎 | 涵盖图数据存储  具备海量图数据存储能力，构建超大规模图存储结构 | 百万亿级点边规模，可管理集群规模超过200节点，PB级存储规模；动态数据更新吞吐量≥1TB/s；冷热数据迁移延迟≤2s/GB. | 280TB，万亿边，[Neo4j 2021](https://neo4j.com/blog/developer/behind-the-scenes-worlds-biggest-graph-database/) https://github.com/neo4j/trillion-graph?tab=readme-ov-file | （十/百）万亿级点边规模，可管理集群规模达百节点，百TB级存储规模 |
| 高可靠的跨域容错恢复机制 | 涵盖容错恢复  可靠性超越国际主流水平 | RPO=0；RTO≤60秒；PB级数据检查点时间≤3分钟；节点故障恢复数据丢失率=0% | 根据中国国家标准《信息系统灾难恢复规范》（GB/T 20988）第 6 级要求，关键业务需达到RPO=0，RTO 分钟级 | RPO=0；RTO≤60秒 |
| 多粒度智能索引体系构建 | 涵盖智能索引 | 动态更新响应时间≤10ms；索引空间占用率≤10%原始图大小 | （智能索引，更低占用率） | （数据量？）分钟级索引构建时间;秒级索引维护时间; 索引空间占用率 < 30 % |
| 超大点分片管理与负载均衡技术 |  | 载荷均衡度≥0.9；多跳查询/图计算的分片间通信总量降低≥30%；动态分片响应时间≤30 s；分片局部性≥85%本地访问率 | Fan et al., "Application Driven Graph Partitioning", SIGMOD 2020  （1%/1%。超大点） | 载荷均衡度≥0.9；（数据量/增量）动态分片响应时间≤30 s； |
| 子课题2：分布式数据库优化器与执行引擎 | 高并发低延迟图查询技术 |  | QPS > 百万；  LDBC FinBench | [LDBC SNB Interactive](https://ldbcouncil.org/benchmarks/snb-interactive/) GraphScope 37000~80000 QPS (using Cypher)  （分布式/单机）  Use FinBench instead ~1000 | QPS> 30000 |
| 分布式多跳查询执行效率的优化方法 | 涵盖多跳查询  全流程多跳查询性能达到国际领先水平 | 多跳查询响应时间≤100 ms（3跳查询）和吞吐量≥1000 QPS；基数估计准确率≥90%；查询执行计划效率提升比≥2×；分布式查询节点间通信总量减少率≥40%  LDBC SNB BI | [秒级查询（6跳查询）](https://www.toutiao.com/article/7336389559858741771/?upstream_biz=doubao&source=m_redirect&wid=1752481546497)  目前主要的可查到的QPS数据都是出自LDBC SNB  (小世界 6跳就全图，mesh) | 多跳查询响应时间≤100 ms（3跳查询）- 1s（6跳查询）；QPS ？ |
| 高性能图计算引擎的细粒度优化 |  | PB级规模图数据在经典图计算任务（如SSSP、BFS等）上的响应延迟BFS ≤ 1s / SSSP ≤ 3s（在10B边规模）与吞吐量≥10M edges/sec；通过Graph500测试；并行可扩展性扩展比 ≥80  Graph500 BFS/SSSP  LDBC graphyalytics | Graph 500 榜单（难度较高）  （其它指标？LDBC Graph Analytics，偏功能完整性，涵盖主流图分析任务） | TB规模图数据（换成点、边数、Scale）在经典图计算任务SSSP、BFS上的GTEPS > 10,000 / 100,000 |
| 子课题3：统一图数据管理和全栈生态融合 | 面向HTAP混合负载的分布式图数据库系统架构 | 海量图数据集的高效处理 | TP/AP数据freshness低于 1s vs. 1min？混合负载资源隔离失效概率≤0.1%；跨负载资源调配响应时间≤5 s （5s->?） | 《GraphScope: A Unified Engine for Large-Scale Graph Processing》（VLDB 2021） | 事务延迟干扰率：TP/AP共享资源池时候的TP平均延迟受干扰程度<15%，AP任务中断率<0.5% |
| 查询语言的智能语义融合 | 整合openCypher和ISO GQL两种主流图查询语言 | 双语言语法兼容性≥95%（openCypher + ISO GQL）；查询自动优化率≥80%（可重写查询比例）；插件API覆盖度≥90%关键函数；生态工具链响应延迟≤200 ms（常用命令） |  |  |
| 图数据库全栈软件生态集成 |  | 支持GUI 能够展示10000点边；导入导出速率 xx MB/s； |  |  |