



CCF-华为胡杨林挑战基金 数据库分项

2023年项目申请指南

1 总则

“CCF-华为胡杨林挑战基金”由CCF和华为联合发起设立，用于支持我国泛计算领域的发展：

- 推动科研围绕产业技术挑战开展，更好实现产业和学术的协同创新；
- 培养具备全球视野，兼具科学研究和解决产业技术问题能力的综合型人才；
- 促进中国泛计算领域的软硬件产业技术创新和生态繁荣；

“CCF-华为胡杨林挑战基金 数据库分项”秉承上述理念，支持在数据库领域围绕产业技术挑战开展研究。

“CCF-华为胡杨林挑战基金”是原“CCF-华为胡杨林基金”新的分支模式，不影响原“CCF-华为胡杨林基金”各专项的运作。

2 申请条件

本基金将面向符合如下条件的国内所有高校及科研院所学者展开：

- 申请者必须是国内高校/科研院所在职的全职教师或研究人员；
- 申请者必须是CCF会员；
- 申请者具有博士学位或者研究生指导资格；
- 申请者如果已经承担2022年“CCF-华为胡杨林基金-数据库专项”项目，在申请本挑战基金时应注明投入资源同“CCF-华为胡杨林基金-数据库专项”项目无重叠。



3 申请课题

支持数据管理方向的创新型基础研究工作，本年度挑战课题如下：

一 数据库轻量化自适应基数估计算法

构建算力要求低，资源消耗少（轻量化），并且能够随着负载变化快速调整和修正（自适应），估计结果精准的智能基数估计框架和算法。能与数据库内核融合，在复杂查询场景中给出比传统基数估计更准确的结果。

二 数据库语义感知的高效压缩

面向TP场景，设计一种可结合数据库语义的压缩算法，支持单个页面（8KB）的压缩，保持合理的压缩性能（不低于单核100MB/秒），基于TPCH标准数据测试集，实现不低于2:1压缩率，点查场景下，开启压缩与不开启压缩相比，挑战执行时间增加不超过5%。

三 数据库负载自适应缓存页面淘汰算法。

构建高命中率、高性能的自适应页面淘汰算法。能够在数据库的业务场景有变化的情况下，自适应地确定需要淘汰的缓存页面。

- 周期性调整客户端Warehouse ID，平均缓存命中率提高20%（比 Clock-sweep 算法），不降低当前数据访问的速度；
- 随机调整客户端Warehouse ID，平均缓存命中率提高10%（比 Clock-sweep 算法），不降低当前数据访问的速度；

四 存算分离架构缓存数据控制权分配

分布式缓存读写页面尽可能本地读写页面，减少远程读页面的概率。一方面优化分布式缓存的页面控制权分配模式，另一个方面是探索在业务接入层通过事务亲和性路由算法达到业务层按照分区接入效果，提高本地读写页面。TPCC业务随机接入执行相比按照warehouse分区接入执行性能劣化不超过10%。

五 分布式优化器中模型构建和求解算法



建立新的分布式优化器的模型和求解算法，从理论上保证可以找到最优解，

并且算法复杂度跟当前单机优化器的复杂度接近。

六 Auto Analyzer统计信息收集自动化

数据库的查询优化器非常依赖于收集的统计信息(histogram, distinct count, min/max), 然而数据更新后, 有些统计信息会过时, 需要更新。但是重新收集全部统计信息的过程非常耗时。

探索在给定一个数据库、历史查询日志、数据更新日志, 自动统计信息收集技术研究 i) 在什么时候 ii) 应该以何种方式(如: 哪些表, 哪些列, 采样率) iii) 收集哪些统计信息(histogram, distinct count, min/max)可以让数据库总体性能最好。

七 以应用性能与可靠性驱动的资源评估推荐方法

基于特定的openGauss数据库版本、硬件配置和应用负载(如不同运行参数下的TPCC/Sysbench/YCSB等), 获取默认数据库参数下的应用负载的性能(如吞吐量/响应时间)和可靠性指标(如MTTR), 作为推荐基线以特定的应用负载(如不同运行参数下的TPCC/Sysbench/YCSB等)和软硬资源作为评估对象, 基于不同的性能和可靠性目标, 推荐相应的CPU/内存/IO组合以及相应的参数设置, 为用户实现基于成本最优的资源调整提供指导。

八 数据库多租环境下的缓冲区共享问题

设计一种可用于数据库内核多租环境下的共享缓冲区替换算法, 基于上述算法约束, 在固定的收益下, 挑战实现最小化成本。

九 数据仓库混合负载环境下的内存管理与隔离问题

设计一种可用于数据仓库内核混合负载环境下的内存调度算法, 可以在保证不同负载按照内存资源分配配比使用内存资源, 达到内存隔离的效果, 同时最大化系统负载吞吐。实现支撑以上内存调度算法的内存管理框架。



申请课题的详细描述（背景、指标、参考文献等）请参见：《CCF-华为胡杨林挑战基金-数据库分项 课题解析报告》，华为将为课题承接者提供技术环境、数据集，华为会为每个项目配备对口的专家，同项目研究者一起进行协同研究。

4 申请时间点

2023年3月07日 指南发布，申请开始；

2023年4月15日 申请截止，启动评审；

2023年4月28日（预计） 评审截止，发布评审结果；

2023年5月30日 前与CCF签署协议（如有特殊情况可适当延长）；

其他相关时间点安排：

2023年4月 项目开题报告；

2023年9月 进行中期答辩，具体时间和地点待定；

2024年5月 进行结题答辩，具体时间和地点待定。

项目申请书需在截止之日24:00前发给项目组专用邮箱dbfunding@ccf.org.cn。任何针对项目申报的问题也请发给本邮箱咨询。

5 项目评审

该项目依托CCF运作，由胡杨林挑战基金数据库分项技术管理委员会负责监督计划的实施和监督，以及项目的评审。委员会评审时主要考虑：

- 申请项目的应用前景、创新性、可行性；
- 申请者（及团队）的学术水平和科研能力；
- 申请项目和指南的匹配程度；
- 申请项目是否将向中国自有开源数据库系统项目openGauss等提交代码。

经过委员会确认授予资助的研究项目需签署合同生效。



6 资助金额

单项目资助额度不超过80万元人民币。

7 研究期限

项目研究期限一般为一年^{注1:}，对于合作成果较好的项目可以滚动支持。

项目立项后不可更换受资助人。

在项目研究工作中，如因受资助者自身原因中断研究工作，项目终止并根据申报时的具体阶段经费使用说明，退回已拨经费的余额。

获得资助的申报者原则上不可放弃资助，如有特殊情况，需提交《放弃声明》并加盖被资助者所在单位公章后由指导委员会存档留备。

8 交付成果

项目可交付成果但不限于：

- (1) 论文、著作、专利等探索内容；
- (2) 项目报告；
- (3) 项目实施开发基于 openGauss 等的原型和算法系统；
- (4) 项目合作发表的论文及著作需标注“受 CCF-华为胡杨林挑战基金资助（英文：Sponsored by CCF-Huawei Populus euphratica challenge Research Funding）”。

注1： 由于项目不确定性等原因，申请者可以申请最长延期6个月结项。

9 双方义务与知识产权

受资助者在项目研究过程中形成的与项目相关的成果的著作权及专利等，包括但不限于论文、著作、源代码等，其知识产权权利归属申请方及华为公司共同所有。华为公司有权免费优先使用。

具体细节以《CCF-华为胡杨林挑战基金数据库分项合同》为准。



10 项目管理

被资助的项目申请者是项目管理的第一责任者。

华为方项目协调人负责对接华为资源，与项目负责人协商确定沟通交流计划。

项目执行的最终管理者是CCF-华为胡杨林挑战基金数据库分项技术管理委员会。

附录1：CCF-华为胡杨林挑战基金 数据库分项 技术管理委员会名单

	华为技术有限公司	CCF
主任		李战怀
委员	冯柯	李国良
	任阳	钱卫宁
	曾凯	周可
执行秘书	石腾	陈跃国