



第11期
2023/11

拥抱AIGC时代柏睿
向量数据库
性能超 Faiss 20倍

做中国的国际智能数据算力公司

AN INTERNATIONAL DATA + AI COMPUTING POWER
COMPANY FROM CHINA

分布式内存数据库 RapidsDB

产品更新

RPDSQL数据库：

1.在SELECT ... INTO HDFS语法中新增将分区数据小文件合并成一个数据文件的功能。

RapidsDB

1.RapidsDB的管理工具Pilot完成，提供方便和强大的集群安装、管理、运维、监控功能。

2.RapidsDB库内推理方面完成库内推理框架设计。

3.自动化测试工具SETH更新，提供并行执行测试功能，加速测试时间



柏睿智能开发平台 Rapids AI

产品更新

【Data Assistant】

基于大模型和柏睿数据库的智能化数据分析工具，帮助用户通过自然语言进行数据查询和分析

1. 实现基于电力数据的智能问答
2. 对大模型进行微调，提升回答准确率和适用性

【AI4DB】

使用人工智能技术对数据查询优化引擎中的基数进行估计，提升优化器的优化准确率

1. 实现基于查询的基数估计模型
2. 实现数据库内推理的框架搭建

存算一体机HyperCube



- 由北京市经济和信息化局、朝阳区人民政府指导，华为、鲲鹏联创主办，柏睿数据等协办的“山高海阔·共赢先机——鲲鹏一体机技术应用大会”在北京成功举办。会上，柏睿数据与华鲲振宇、鲲鹏联创正式签署战略合作协议，并联合发布整合各方软硬件优势打造的核心产品——“智算一体机”



发展战略合作伙伴

- 正式成立首批鲲鹏一体机CTO技术圈，柏睿数据等共10家鲲鹏生态企业成为首批技术圈成员
- 智算一体机雏形产品方案已在运营商、智能制造、医疗等核心业务场景落地应用

柏睿向量数据库Rapids VectorDB

显著降低AIGC“入门”成本，柏睿向量数据库助力打造高效易用的向量数据引擎

- Rapids VectorDB 在检索过程中无需使用 CPU / GPU，即可实现比在 CPU / GPU 上运行的Faiss高 20+ 倍的整体性能

测试拓扑



配置环境

	主机类型	主机配置	GPU
柏睿向量数据库	虚拟机	计算节点: 8vCPU,16G内存 *1 存储节点: 8vCPU,32G内存 *3	无
专用向量数据库	物理机	CPU: 32线程(IntelE5-2620 v4 2.10GHz*2) 128G内存 *1	NVIDIA RTX 3060(12G显存)

测试结果

柏睿向量数据库	全内存分布式查询: 0.99秒		
专用向量数据库 (单机CPU)	磁盘->内存加载数据: 21.19秒	CPU查询: 2.1秒	
专用向量数据库 (单机GPU)	磁盘->内存加载数据: 21.19秒	内存->GPU 加载: 1.55秒	GPU查询: 0.35秒

根据性能测试结果可知，**Rapids VectorDB 仅需0.99秒**即可返回查询结果，而运行在 CPU 上的专用向量数据库则用**23.29秒**返回结果，运行在 GPU 上的专用向量数据库则需**23.09秒**。可见，Rapids VectorDB 在向量数据检索查询方面不仅具备显著的性能优势，还具有极高性价比。

柏睿数据

产品系列月刊



柏睿数据

北京柏睿数据技术股份有限公司

Borui Data Technology (Beijing) Co. Ltd.

北京市朝阳区利泽西街6号院东湖国际中心A座7层