

# 向量启航，引擎加持

2022年10月

中国数据库行业分析报告





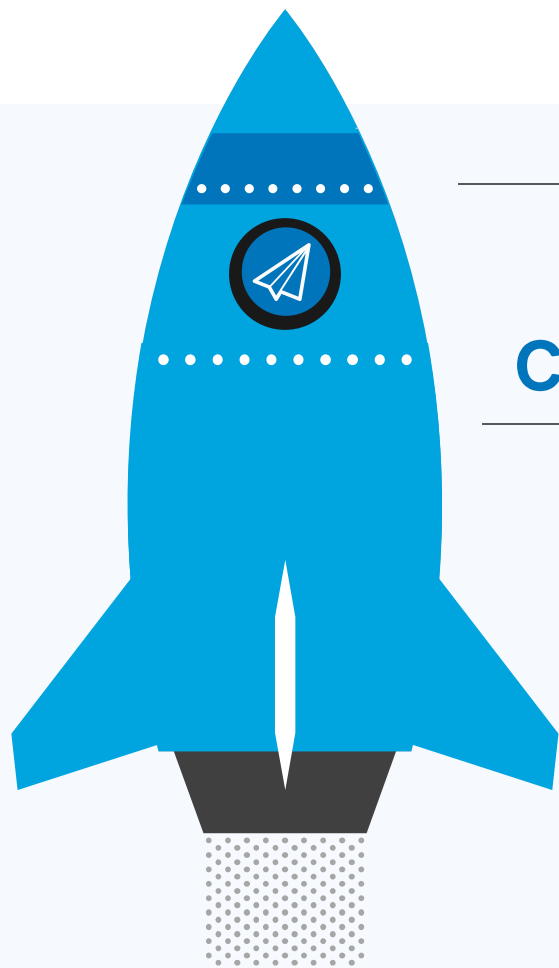
2022年10月的 墨天轮中国数据库流行度排行榜火热出炉，本月共有245个数据库参与排名，相比上月新增七个数据库，本月排行榜前十名变动较大：达梦数据库上市程序恢复，热度大涨，排名反超openGauss，重回第三；华为旗下两大数据库品牌名次均下降；云原生数据库PolarDB、TDSQL再创佳绩，名次均上升一位。本月排行榜从第十一名至第三十名，竞争激烈，归属变动较大。在这一赛道中，本月也崛起了一批数据库新秀，拥有亮眼的表现。



本月报告中墨天轮发布了**最新向量数据库全球产业图谱**，国内典型向量数据库代表有Milvus、Vearch、TensorDB、Om-iBASE等。向量数据库是专门为处理向量嵌入独特结构而构建的数据库系统。它们通过比较值并找到彼此最相似的向量来索引向量，以便于搜索和检索。从技术角度来讲，**向量数据库主要解决2个问题，一个是高效的检索，另一个是高效的分析**。向量数据库其实就像传统数据去处理一些关系型数据、结构化数据一样，承担的是非结构化数据的低成本存储和高性能计算两大核心能力。具体包括用于搜索和检索的向量索引、单级过滤、数据分片、复制、混合存储以及API功能。向量数据库主要的应用领域如**人脸识别、推荐系统、图片搜索、视频指纹、语音处理、自然语言处理、文件搜索**等。随着 AI 技术的广泛应用，以及数据规模的不断增长，向量检索也逐渐成了 AI 技术链路中不可或缺的一环，更是对传统搜索技术的补充，并且具备多模态搜索的能力。



随着数据库软硬件技术的发展，经典的SQL计算引擎逐渐成为数据库系统的性能瓶颈，尤其是对于涉及到大量计算的OLAP场景。如何充分发挥底层硬件的能力，提升数据库系统的性能，成为近年来数据库领域的热门研究方向，而**向量化执行就是解决上述问题的一种有效手段**。火山模型的诞生为缓存数据库的内存压力，但该设计并未充分利用CPU的执行效率且以往的火山模型一次处理一个元组的方式造成过大的解释执行代价，阻止了对性能影响极大的编译优化。2005年《MonetDB/X100: Hyper-Pipelining Query Execution》的论文首次提出“向量化引擎”的概念，后续国产数据库陆续推出向量化执行引擎，**加速OLAP场景的查询分析速度**。



## 目录 content

一、数据库排行榜及前沿动态

二、向量数据库的概述与解析

三、向量引擎加持传统数据库

四、向量数据库产品相关案例

# 实干兴邦 - 前四强合成 TODO 促成长



秋风萧瑟，洪波涌起。2022年10月的墨天轮中国数据库流行度排行榜火热出炉，本月共有245个数据库参与排名，相比上月新增七个数据库，本月排行榜前五名变动较大：达梦数据库上市程序恢复，热度大涨，排名反超openGauss，重回第三；PolarDB上升一位居第五。

2022年10月中国数据库排行榜TOP5

排行	上月	半年前	名称	模型	属性	三方评测	生态	专利	论文	得分	上月	半年前
	1	1	TiDB +	关系型	HP X B C			15	23	612.45	+31.50	-4.39
	2	↑ 3	OceanBase +	关系型	HP X B C			137	17	584.12	+47.40	+40.90
	↑ 4	↑↑ 5	达梦 +	关系型	TP X B C			381	0	556.12	+20.90	+100.50
4	↓ 3	↓↓ 2	openGauss +	关系型	TP X B C			562	65	533.33	-2.37	-34.09
5	↑ 6	↑ 6	PolarDB +	关系型	X B C HP			512	26	436.30	-5.08	+48.47

PolarDB作为国产云原生数据库中的佼佼者，本月排名赶超GaussDB这一云上竞争对手。其本月得分下降5.08分，以不到一分的微弱优势排名第五。

TiDB本月得分612.45分，较上月得分环比上涨5.4%。这也是其今年6月重夺榜单第一宝座后，连续五个月稳坐榜首。自2020年1月至今，TiDB已累计霸榜33个月。

OceanBase 本月得分 584.12分，与第一名得分差距从上个月的44.23分缩小至28.33分。其本月分数涨幅最大，较上月得分上涨8.8%，排名第二。

达梦本月得分较上月上涨20.9分，以556.12分摘得探花。其上月排名被反超后，一直奋力向上，本月排名赶超openGauss。达梦自递交入股申请书后，一直备受关注。

openGauss本月得分较上月仅下降2.37分，排名却下降一位居第四。9月30日，openGauss迎来了里程碑事件，openGauss3.1.0版本正式上线，此版本与之前版本特性功能保持兼容。

# 稳中求进 - 金仓、TDSQL立足创新创佳绩



2022年10月中国数据库排行榜TOP6-TOP10

排行	上月	半年前	名称	模型	属性	三方评测	生态	专利	论文	得分	上月	半年前
6	↓ 5	↓↓ 4	GaussDB +	关系型	☒ ☒ ☒	☒ ☒	☒ ☒ ☒ ☒	562	65	435.39	-37.88	-40.18
7	7	↑ 8	人大金仓 +	关系型	TP ☒ ☒	☒	☒ ☒ ☒ ☒ ☒	232	0	431.37	+12.90	+108.69
8	↑ 9	↑ 9	TDSQL +	关系型	☒ ☒ ☒ ☒	☒ ☒	☒ ☒ ☒ ☒	39	10	279.70	+2.44	+8.77
9	↓ 8	↓↓ 7	GBase +	关系型	TP ☒ ☒	☒ ☒ ☒	☒ ☒ ☒ ☒ ☒	152	0	275.17	-10.70	-97.25
10	10	10	AnalyticDB +	关系型	AP ☒ ☒	☒ ☒ ☒	☒	480	28	192.52	-13.64	+3.32

## AnalyticDB

本月得分较上月下降13.64分，连续13个月蝉联墨天轮排行榜第十名。其是阿里云自主研发的一款实时分析数据库，在云上拥有一席之地。新环境下，AnalyticDB也一直在打磨产品。

## GaussDB

其是华为云自研数据库的统一品牌，本月得分435.39分,以不到1分的微弱劣势被反超。9月，GaussDB亮相华为全联接2022·曼谷站，其动向不太频繁，热度上有所降低。

## 人大金仓

其是成立最早的国产数据库厂商，据太极股份的半年财报，人大金仓2022上半年营收1.23亿、净利润940万。其本月以4.02分的分数劣势，排名第七。

## TDSQL

其是腾讯云企业级分布式数据库，本月得分279.70分，以4.53分的优势领先GBase。近日，腾讯云数据库以其过硬的产品，成功中标中国邮政4年订单。

## GBase

其是南大通用数据技术有限公司推出的自主品牌的数据产品。九月，GBase南大通用数据库相继中标成都农商行&自贡银行&泉州银行等多个重点项目。



# 异军突起 - 后起之秀细分领域闪耀光芒



本月排行榜从第十一名至第三十名，竞争激烈，归属变动较大。在这一赛道中，本月也崛起了一批数据库新秀，拥有亮眼的表现。

 2022年10月中国数据库新秀得分详情表

排行	上月	半年前	名称	模型	属性	三方评测	生态	专利	论文	得分	上月	半年前
17	↑↑ 19	↑↑↑ 29	MogDB +	关系型	TP H			27	0	62.61	+5.77	+30.78
20	↑↑ 22	↑↑↑ 23	DolphinDB +	时序	X			0	0	55.21	+5.30	+16.00
21	↑↑↑ 27	↑↑↑ 44	StarRocks +	关系型	AP B X			2	0	54.54	+14.33	+32.43
27	↑↑ 29	↓↓↓ 19	TGDB +	图	C X		-	0	0	38.30	+3.78	-6.03
28	↑↑↑ 31	↑↑↑ 31	CTSDB +	时序	X		-	0	0	36.31	+3.34	+5.83
30	30	↓↓↓ 21	KunDB +	关系型	TP X		-	83	0	33.43	+0.02	-9.65

**CTSDB**  
墨天轮排行榜上时序数据库第三名CTSDB，在整体排名中较上月排名上升三位至第28名。CTSDB是腾讯唯一的时序数据库，其支撑了腾讯内部20多个核心业务。

**KunDB**  
其是2019年星环科技推出了一款分布式关系型数据库，其本月排名较上月虽未发生变化，但是实力不容小觑。近日，星环科技获得证监会批准，正式进入科创板IPO发行阶段，将成为“国产大数据基础软件第一股”。

**MogDB**  
云和恩墨基于 openGauss 内核进行增强提升，推出的一款安稳易用的企业级关系型数据库MogDB，本月排名上升两位至第17名，逐渐逼近前十赛道。上个月MogDB力争上游，在市场拓展和生态建设上都卓有成效。

**StarRocks**  
北京鼎石纵横科技有限公司于2020年推出的一款新一代极速MPP分析型数据库系统，本月排名跃升六位至第21名。9月24日，年度盛典 StarRocks Summit Asia 2022 顺利举行，9月27日，StarRocks2.3.3重磅发布。

**DolphinDB**  
由浙江智臬科技有限公司研发的一款高性能分布式时序数据库,公司主创团队从2012年开始投入研发，本月排名上升两位至第20名，也是排行榜上排名第二的时序数据库。



**TGDB**  
腾讯云推出的原生分布式并行图数据库TGDB是排行榜上图数据库第一名。其排名上升两位至第27名。它不仅具备图数据库的优点，还兼具原生图数据库的关联关系深链查询能力和分布式图数据库的数据延展性及计算性能。

# 产品动态 - openGauss 3.1.0版本正式发布



2022年9月30日，openGauss 3.1.0版本正式上线！openGauss 3.1.0版本是 openGauss 2022年发布的Preview版本，版本维护生命周期为半年。此次发布包含两个数据库服务端安装包：企业版和轻量版。

openGauss 3.1.0版本与之前版本特性功能保持兼容，在**企业级特性**、**高可用**、**高性能**、**高智能**、**高安全**、**工具链**、**可扩展性**七大特性上全面增强。

## 企业级特性

1. 行存表压缩能力增强
2. 发布订阅能力增强
3. 细粒度滚动升级
4. statement\_history视图诊断能力增强

## 高可用

1. 两地三中心跨Region容灾
2. CM支持对外状态查询和推送能力
3. DCF (Distributed Consensus Framework, 分布式共识框架, 基于Paxos算法实现数据同步强一致。)支持策略化多数派

## 高性能

### 基础算子性能提升

- 新选择率模型典型场景选择率估算准确率、性能提升1X
- 分区表页面估算优化典型场景性能提升20%。
- Partition Iterator算子优化典型场景性能提升5%。
- 函数依赖特性支撑多列查询典型场景行数估算准确率提升1X。

## 可扩展性

集成openLookeng, 提供分布式OLAP能力

基于openLookeng实现分布式分析能力，openLookeng复用ShardingSphere中间件的分库分表能力，使openLookeng可以获取openGauss数据进行分析运算。加上ShardingSphere搭配openGauss形成的分布式OLTP能力一起组合成分布式的HTAP能力。

[立即体验:](https://opengauss.org/zh/download.html)

<https://opengauss.org/zh/download.html>

## 高智能

1. DBMind自治运维平台  
构建端到端自治运维平台：新增异常检测能力，完善自监控、自诊断、自调优能力。
- 2、智能优化器
  - 实现库内Bayes网络算法并基于此实现智能统计信息以提高多列基数估计准确度。
  - 计划自适应选择解决因数据倾斜等跳变难题。

## 高安全

### 细粒度Any权限增强

Any权限管理，新增支持5种对象共12种：

- ALTER ANY TYPE、DROP ANY TYPE
- ALTER ANY SEQUENCE、DROP ANY SEQUENCE、SELECT ANY SEQUENCE
- ALTER ANY INDEX、DROP ANY INDEX
- CREATE ANY TRIGGER、ALTER ANY TRIGGER、DROP ANY TRIGGER
- CREATE ANY SYNONYM、DROP ANY SYNONYM

## 工具链

1. MySQL全量迁移性能提升
2. MySQL增量迁移支持事务级并行消费，提升增量迁移性能
3. 支持基于默克尔树的数据校验
4. 支持openGauss到MySQL迁移，满足MySQL反向迁移要求

# 产品动态 - Oracle 23c新特性和发布周期计划

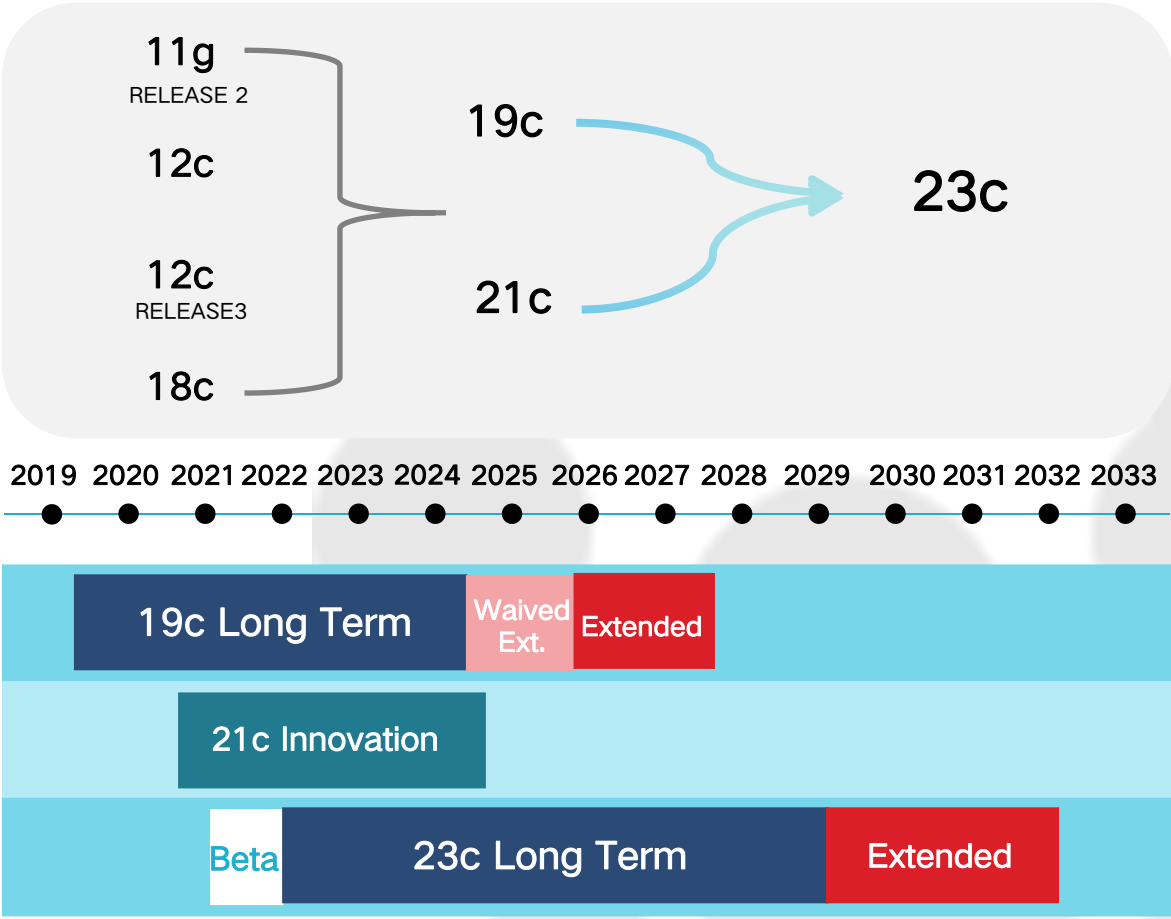


DOAG 2022 会议上，Oracle 的产品经理 Gerald Venzl 披露了 Oracle Database 23c 的一系列新特性。Oracle Database 23c 的发布计划已经明确公布，在2022年，Beta版已经开始测试。新版本将在2023年发布，23c 是一个长期支持版本。

## Oracle 23c 十小新特性

- 01 不带FROM子句的SELECT查询
- 02 单表支持4096列
- 03 SCHEMA 级别的权限
- 04 Boolean 数据类型
- 05 基于别名和位置的GROUP BY
- 06 Javascript 存储过程
- 07 SQL Domains
- 08 DDL的 IF EXISTS判断
- 09 数据库对象增加注释
- 10 标准的表值构建

## 升级到Oracle 23c的路径

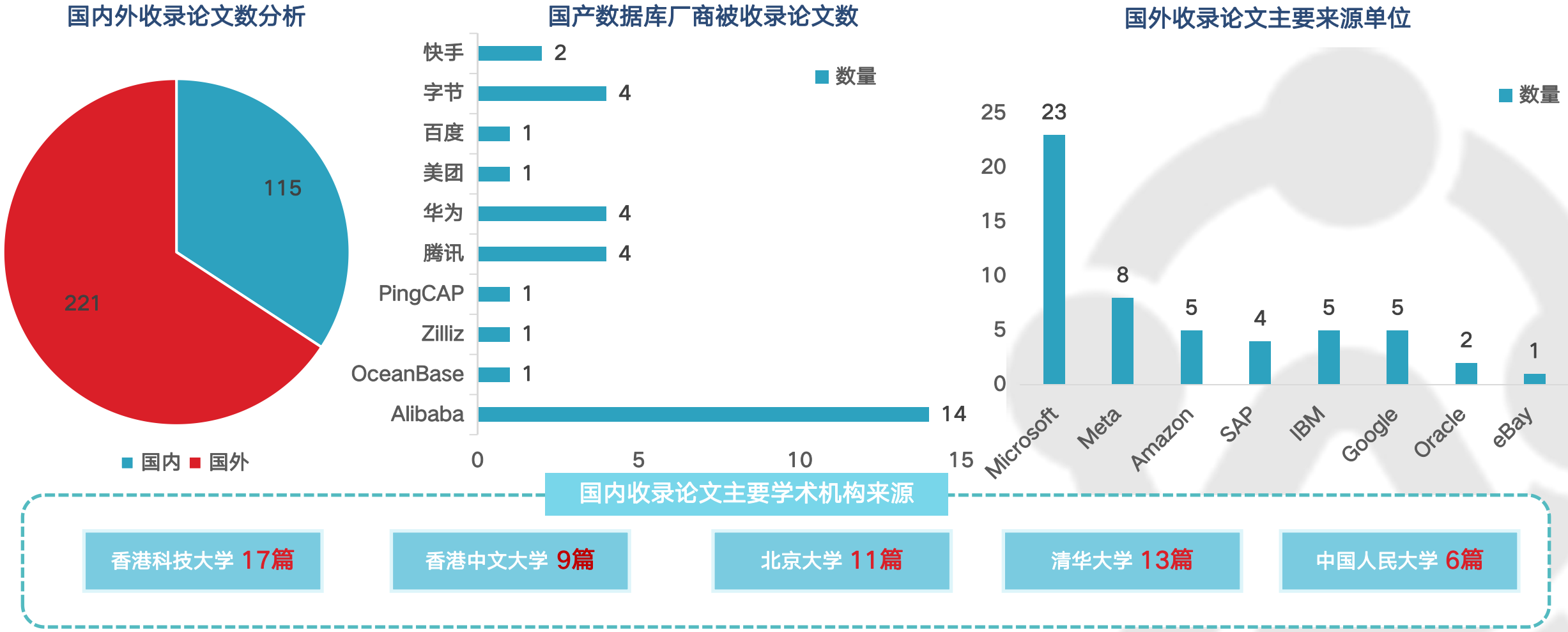




# 学术动态 - 中国在VLDB2022的论文盘点



VLDB2022于9月5日召开，VLDB（Very Large Data Base）作为数据库领域的三大顶级国际会议之一，是面向数据库研究人员，内核开发人员，开发商以及用户的年度国际会议论坛，代表数据库系统领域最杰出的研究和工程进展。VLDB2022会议中共有336篇国内外论文入选，其中中国贡献115篇，占比超过1/3。由于单篇论文有多个作者，来源地不同，以下数据含重复计数。



# 调研动态 - 四家图数据库厂商入选Gartner调研报告



近日，国际知名调研机构Gartner发布了聚焦图技术的调研报告——《图数据库管理系统市场指南》（以下简称“指南”），在全球范围内，甄选出32家图数据库代表性供应商，Galaxybase、AtlasGraph、Ultipa、StellarDB四个数据库作为优质图数据库入选《指南》，获得了业界积极评价和高度认可。《指南》从图技术市场现状，未来发展方向、图数据库选型等多个维度深入分析，明确市场发展趋势和竞争格局，为企业客户提供战略参考。

## 《图数据库管理系统市场指南》亮点

01

### 图数据库市场趋势性预测

- 到2025年，包括图数据库管理系统(DBMSs)在内的图技术市场将增长到32亿美元，年复合增长率为28.1%”
- 到2025年，图技术将用于80%的数据和分析创新，高于2021年的10%，促进企业快速决策”。
- 大型传统数据库管理系统和平台供应商以及初创公司都在瞄准机会。

02

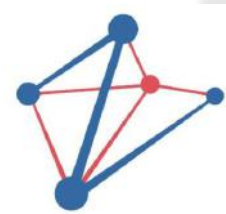
### 图数据库受众画像

- 程序开发员：正在将更多地面向客户或内部的项目转向图技术，利用图数据库作为存储和执行后端。
- 数据架构师：正在为内容管理、个性化和语义数据兼容性设计基于知识图谱的解决方案。
- 数据科学家：对数据点、边之间的连接和关系进行高阶探索。
- 业务所有者和领域专家。

03

### 图数据库分类

- Gartner依据底层存储将图数据库大致划分为原生图数据库和多模图数据库两大类型。原生图数据库，相较于多模数据库，更适用于涉及实时计算、多跳查询和机器学习(ML)等资源密集型处理场景。
- 原生图数据库在处理超大规模图(通常是数十亿个节点)的查询时能提供更优质的性能。”



# 评测动态 - 最新电信行业数据库能力测评榜单发布



2022年9月1日，北京软件和信息服务业协会对《电信行业数据库适配测试技术规范》、《电信行业数据库适配测试业务规范》进行了团体标准立项的专家评审。经研判讨论，两项标准正式获批立项。国家工业信息安全发展研究中心依托两项标准组织开展了多轮电信行业数据库能力测评，最新一批入围“场景榜单”的综合排名的前三名：**中兴通讯、亚信科技、阿里云**。

综合排名					
排名	厂商	产品名称	版本	发布时间	技术平台
1	ZTE中兴	GoldenDB	V5	2021.11.20	鲲鹏920+麒麟V10
2	AsialInfo 亚信科技	亚信安慧AntDB数据库系统	V6.2	2021.8.30	鲲鹏920+统信V20
3	阿里云	PolarDB数据库管理软件	V2.0	2019.8.1	鲲鹏920+麒麟V10

《电信行业数据库适配测试技术规范》、《电信行业数据库适配测试业务规范》旨在为电信行业数据库产品能力测评提供依据，以真实业务场景全面验证数据库产品支撑电信级应用的能力，为相关单位测试、选型工作提供参考，推进数据库产品在电信行业的应用推广。

单项排名					
应用场景	评价指标	测试结果	排名1	排名2	排名3
数据加载	最佳成绩 (条/秒)	605941	ZTE中兴	OCEANBASE	阿里云
	平均成绩 (条/秒)	117348			
开户场景	TPS最佳成绩 (事务数/秒)	43955	AsialInfo 亚信科技	阿里云	柏睿数据
	TPS平均成绩 (事务数/秒)	11252			
导入导出	导入最佳成绩 (条/秒)	387599			
	平均成绩 (条/秒)	208238	OCEANBASE	ZTE中兴	GBASE 南大通用
	导出最佳成绩 (条/秒)	2471638			
	平均成绩 (条/秒)	871826			
话单处理	话单生成最佳成绩 (条/秒)	510357			
	平均成绩 (条/秒)	60904	ZTE中兴	腾讯云	AsialInfo 亚信科技
	话单处理最佳成绩 (条/秒)	152337			
	平均成绩 (条/秒)	22937			
话单查询	最佳成绩 (事务数/秒)	118831	阿里云	AsialInfo 亚信科技	天翼云
	平均成绩 (事务数/秒)	41031			

# 商业动态 - 九月国产数据库厂商中标一览



2022年9月国产数据库厂商中标一览表				
公告时间	项目名称	中标数据库	中标金额（元）	采购单位
2022/9/5	某直辖市档案馆数字化运维项目	AntDB	/	某直辖市档案馆
2022/9/13	中移（杭州）信息技术有限公司2022年国产分布式数据库技术服务采购项目	云树系列产品	/	杭州移动
2022/9/13	正数网络2022-2023年数据库产品及技术支撑服务集中采购项目（河南省大数据中心）	CirroData	/	正数网络
2022/9/15	中移动信息2022-2023年分布式OLTP数据库及工具框架采购项目	GodenDB、OceanBase AntDB、GreatDB	共计1.45亿	中国移动
2022/9/19	中原银行2022年信息技术应用创新-OceanBase数据库软件许可采购项目	OceanBase	627 万	中原银行
2022/9/22	中国移动四川公司2021年业务支撑BOSS扩容改造工程国产分布式数据库项目	GreatDB	188.145万	四川移动
2022/9/27	泉州银行新一代智慧审计平台配套设备及数据库采购项目	GBase 8a	/	泉州银行
2022/9/29	正数网络2022-2023年数据库产品及技术支撑服务集中采购项目	UXDB	/	正数网络
2022/9/29	中国邮政技术中台国产关系型数据库和数据备份软件采购项目	TDSQL	/	中国邮政
2022/9/30	2022年第三季度中央国家机关政府采购中心正版软件采购	达梦	48万	国家自然科学基金委员会



# 融资动态 - 时序厂商Greptime完成天使轮融资



9月28日消息，时序数据库厂商 Greptime（格睿云）宣布完成数百万美元天使轮融资，本轮由耀途资本领投，九合创投跟投。Greptime 公司当前正在打磨时序数据库 Greptime DB，未来也计划推出基于Greptime DB的全托管数据库服务Greptime Cloud。



## 公司简介

- 成立于2022年4月，是一家时序数据库厂商。
- 公司产品主要分为 Greptime DB 和 Greptime Cloud。
- 团队方面，当前Greptime员工人数有15人左右，在北京、杭州分设办公室。其创始团队具备在国内互联网大厂从事超大规模监控系统和车联网云平台研发的经验，解决过超大规模混合云架构下的系统运维和监控问题。
- 在计划中，Greptime的客户画像会分为监控领域（可观测）、IoT（智能制造，车联网）和金融三类。
- Greptime会先持续打磨产品，并通过开源的方式持续观察商业化可能，计划在2022年年底将分布式版本开源。2023年初，公司计划推出基于Greptime DB的数据库云服务。

## GreptimeDB 简介

**简介：** Greptime DB 是 Greptime（格睿云）研发打磨的时序数据库产品。

### 特点：

- 用 Rust 编写，可持续且安全
- 集群版本开源，随用随扩展
- 支持 Python 和 SQL
- 亚秒分析查询
- 与现有数据堆栈良好集成



## Greptime云

**简介：** 由 Greptime 完全托管，提供弹性且经济高效的 GreptimeDB 服务轻松快速的配置。

### 特点：

- 协作
- 从各种来源即时获取数据
- 部署在多云上
- 免费升级、备份和安全修复



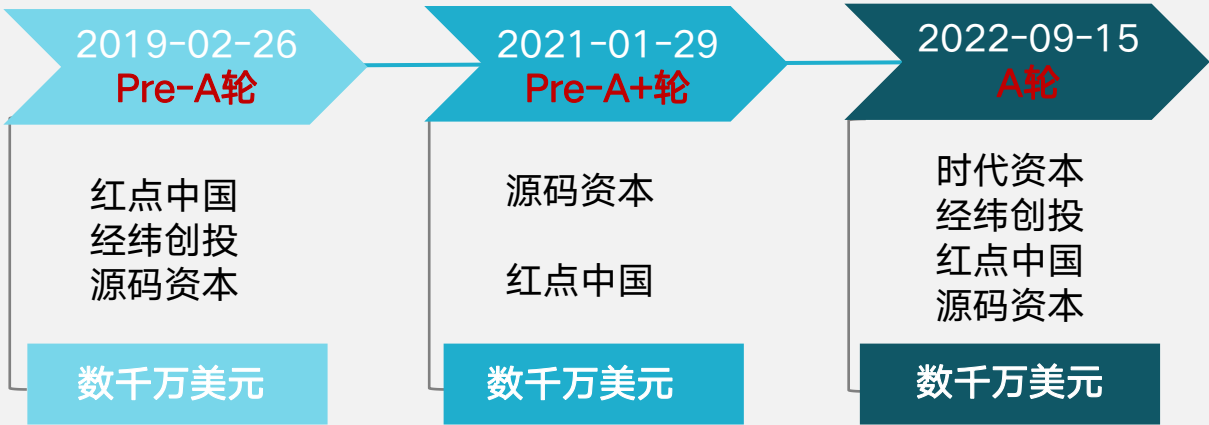


# 融资动态 - Nebula Graph获数千万美元A轮融资



9月15日消息，国内知名的图数据库Nebula Graph研发商杭州悦数科技有限公司宣布获得数千万美元的A轮融资——由时代资本(Jeneration Capital)领投，老股东经纬创投、红点中国、源码资本全部继续加码；华兴资本担任此轮融资独家财务顾问。

## 悦数科技融资历程



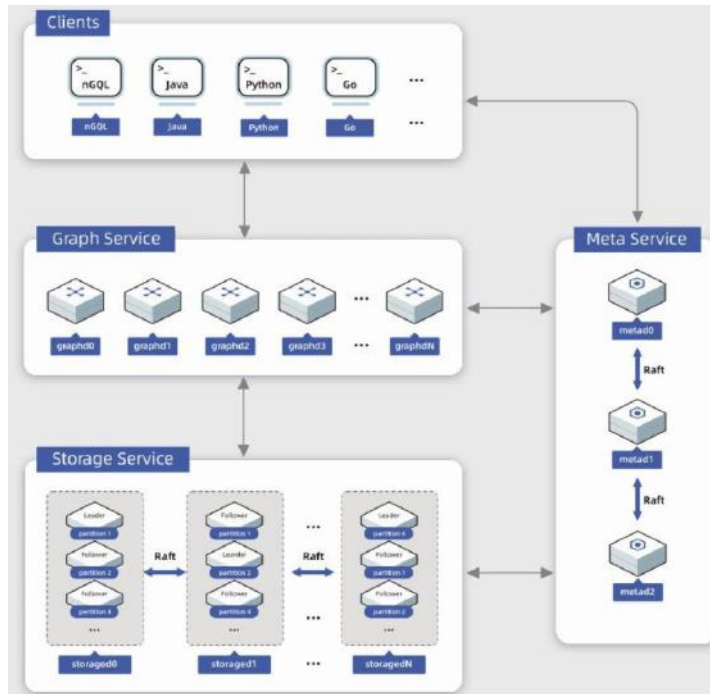
## 悦数科技简介



- 悦数科技成立于2018年10月，是一家科技型创业公司。公司创始团队来自于Facebook、阿里巴巴、华为等国内外各大知名公司。我们致力于世界上唯一开源的分布式图数据库星云的研发，为客户提供稳定高效的互联网基础技术服务。
- 主要产品：**星云图数据库（Nebula Graph）、WEB GUI 工具、图探索可视化分析工具、图数据库可视化监控工具。

## Nebula Graph 产品简介

**1、产品简介：**Nebula Graph 是一款开源分布式图数据库产品，擅长处理千亿节点万亿条边的超大数据集，同时保持毫秒级查询延时，得益于其 shared-nothing 以及存储与计算分离的架构设计，Nebula Graph 具备在线水平扩缩容能力。



## 2、核心特性：

- 自主研发可控，开放系统生态
- 权限控制管理，保障业务安全
- 分布式高可用，保证业务连续
- 实时查询性能，加快业务响应
- 多维扩展能力，助力业务增长
- 复杂查询逻辑，适配业务逻辑

---

## 一、数据库排行榜及前沿动态

---

---

## 二、向量数据库的概述与解析

---

---

## 三、向量引擎加持传统数据库

---

---

## 四、向量数据库产品相关案例

---

向量数据库是专门为处理向量嵌入( vector embedding ) 独特结构而构建的数据库系统。它们通过比较值并找到彼此最相似的向量来索引,以便于搜索和分析。国内典型向量数据库代表有Milvus 、 Vearch、TensorDB 、Om-iBASE等。

## 向量数据库的特征

- 提供标准的 SQL 访问接口,降低用户的使用门槛。
- 提供高效的数据组织,检索和分析的能力。一般用户在存储和检索向量的同时,还需要管理结构化的数据,即支持传统数据库对结构化数据的管理能力。

## 向量数据库的应用

从技术角度来讲,向量数据库主要解决2个问题,一个是高效的检索,另一个是高效的分析。

- 1) 检索通常就是图片检索图片,例如人脸检索,人体检索,和车辆检索,以及猫厂的商品图片检索,人脸支付。
- 2) 分析在平安城市应用的比较多,例如人脸撞库,公安会把2个类似作案手法的案发现场周边的人像做对比,看哪些人同时在2个案发现场出现。

## 向量数据库的关键技术

### 1.构建在大数据和分布式数据库技术基础上

- ✓ 必定是shared-nothing架构
- ✓ 高可用
- ✓ 支持线性扩展

### 2.向量索引技术

- ✓ 向量索引发展,和各种技术的局限性 (LSH,k-d tree, PQ, PQ Fast Scan)
- ✓ 向量与结构化数据的结合

### 3.硬件加速

- ✓ 各种加速硬件的原理,特点
- ✓ FPGA/GPU/AI芯片加速

## 向量数据库与传统数据库的区别

### ➤ 数据规模超过传统的关系型数据库

传统的关系型数据库管理1亿条数据已经是拥有很大的业务流量,而在向量数据库需求中,一张表千亿数据是底线,并且原始的向量通常比较大,例如512个float=2k,千亿数据需要保存的向量就需要200T的存储空间(不算多副本),单机显然不具备这种能力,可线性扩展的分布式系统才是正确的道路,这对系统的可扩展性,可靠性,低成本提出非常大的挑战。

### ➤ 查询方式不同,计算密集型

传统的数据库查询通常可以归结为点查和范围查,而无论是点查和范围查都是一种精确查找,即查询得到的结果要么符合条件要么不符合条件,而向量数据库的向量查询通常是近似查找,即查找与查询条件相近的结果,即查询得到的结果是输入条件最相似的,而近似比较对计算能力要求非常高。

### ➤ 低时延与高并发

在平安城市中的应用需要支持交互式查询,端到端3秒,对向量数据库的要求提升到1秒,我们的设想是后续所有的警察人手一个查询终端,所以高并发也是必须的,1w QPS是我们的底线。

# 向量数据库的核心能力



向量数据库其实就像传统数据去处理一些关系型数据、结构化数据一样，承担的是非结构化数据的低成本存储和高性能计算两大核心能力。具体包括用于搜索和检索的向量索引、单级过滤、数据分片、复制、混合存储以及API功能。

## 用于搜索和检索的向量索引

向量数据库使用专门算法来有效地索引和检索向量。不同的用例需要优先考虑准确性、延迟或内存使用，可以使用不同的算法进行微调。除了索引之外，还有相似度和距离指标，用于衡量向量之间的相关性/相似性。向量索引的常见指标包括欧氏距离、余弦相似度和点积。向量数据库使用“近邻（NN）”索引来评估对象之间或与搜索查询之间的相似程度。传统的近邻搜索对于大型索引来说是有问题的，因为它们需要在搜索查询和每个索引的向量之间进行比较。比较每个向量需要时间。近似近邻（ANN）搜索通过近似和检索最相似向量的最佳猜测来规避这个问题。虽然ANN不能保证返回准确的最接近的匹配，但它在精度和速度之间取得了平衡。

## 单级过滤

筛选允许根据向量元数据来限制搜索结果。可以通过返回基于限制标准的可用匹配子集来提高搜索结果的相关性。后期过滤首先应用近似近邻搜索，然后将结果限制在元数据过滤限制上。用元数据对向量进行预过滤可以缩小数据集，并可能返回高度相关的结果。然而，由于预过滤首先对索引中的每个向量应用匹配标准，它也会严重降低向量数据库的性能。单级过滤结合了预过滤的准确性和相关性，其速度与后过滤一样快或更快。通过将向量和元数据索引合并为一个索引，将两种方法结合起来以达到最佳效果。

## 数据分片

ANN算法可以高效搜索向量。但无论其效率如何，硬件限制了向量在单台机器上的可能性。将向量划分为碎片和副本，在许多商品级机器上进行扩展，以实现可扩展性和具有成本效益。向量数据库将向量平均分成碎片，搜索每个碎片，并在最后将所有碎片的结果结合起来，以确定最佳匹配。通常，使用Kubernetes，并授予每个分片自己的Kubernetes pod，至少有一个CPU和一些内存。这些pod可并行搜索向量。

## 复制

分片允许向量数据库采用许多pods以并行的方式来更快地执行向量搜索。但是，如果需要同时或快速连续地执行许多不同的向量搜索呢？复本复制了整个pod集，以并行处理更多的请求。复本还可以提高可用性。向量数据库可以将副本分散到不同的可用区，以确保高可用性。

## 混合存储

使用混合存储，压缩的向量索引存储在内存中，完整的向量索引存储在磁盘上。内存索引可以将搜索空间缩小到磁盘上全分辨率索引内的一小组候选项。混合存储允许企业在相同的数据占用空间中存储更多向量，通过提高整体存储容量来降低运行向量数据库的成本，而不会对数据库性能产生负面影响。

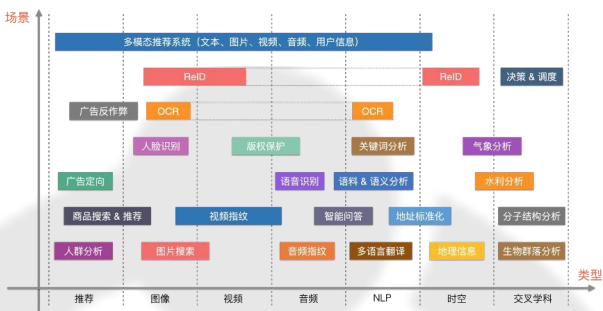
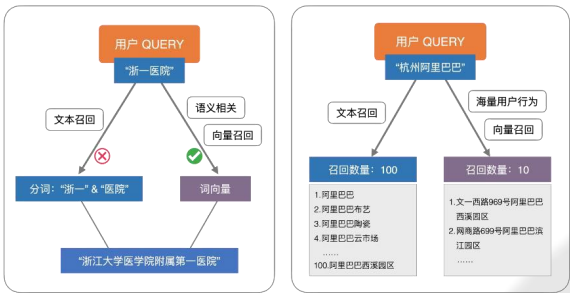
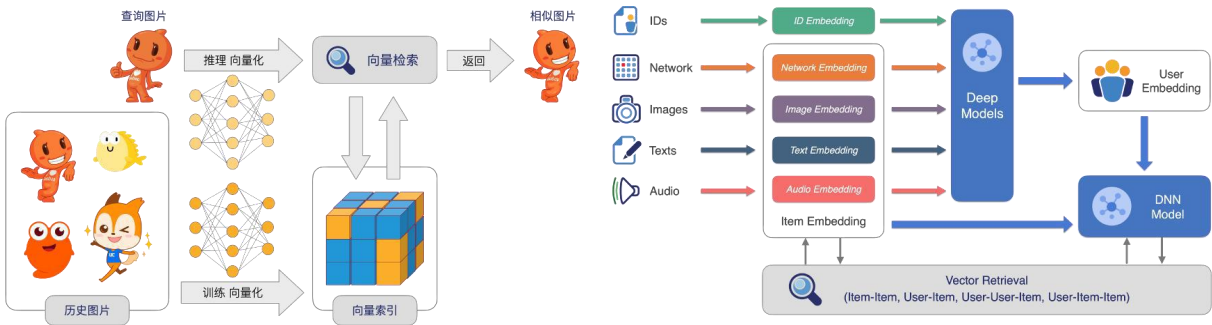
## API

与许多托管服务一样，应用程序通常通过API与向量数据库进行交互。这使企业可以专注于自己的应用程序，而不必担心管理自己的向量数据库的性能、安全性和可用性挑战。API调用使开发人员和应用程序可以轻松上传、查询、获取结果或删除数据。

# 向量数据库应用场景



向量数据库主要的应用领域如人脸识别、推荐系统、图片搜索、视频指纹、语音处理、自然语言处理、文件搜索等。随着 AI 技术的广泛应用，以及数据规模的不断增长，向量检索也逐渐成了 AI 技术链路中不可或缺的一环，更是对传统搜索技术的补充，并且具备多模态搜索的能力。



## 语音、图像、视频检索

向量检索的第一大类应用就是对语音、图像、视频这些人类所接触到的，也最为常见的非结构化数据的检索。

以图片搜索为例，先以离线的方式对所有历史图片进行机器学习分析，将每一幅图片抽象成高维向量特征，然后将所有特征构建成高效的向量索引，当一个新查询（图片）来的时候，对其进行分析并产生一个表征向量，然后用这个向量在之前构建的向量索引中查找出最相似的结果，这样就完成了一次以图片内容为基础的图像检索。

## 搜索、推荐、广告

在电商领域的搜索/推荐/广告业务场景中，常见的需求是找到相似的同款商品和推荐给用户感兴趣的商品，这种需求绝大多数都是采用商品协同和用户协同的策略来完成的。新一代的搜索推荐系统吸纳了深度学习的 Embedding 的能力，通过诸如 Item-Item (i2i)、User-Item (u2i)、User-User-Item (u2u2i)、User2Item2Item (u2i2i) 等向量召回的方式实现快速检索。

## 文本检索

上左图以搜索“浙一医院”为例，如果使用文本分词“浙一”和“医院”，是搜索不到结果的。如果能够利用人们对人们历史语言，甚至历史的点击关联进行分析，建立起语义相关性的模型，把所有的地址都用高维特征来表达，那么“浙一医院”和“浙江大学医学院附属第一医院”的相似度可能会非常高，因此可以被检索出来。上右图以搜索“杭州阿里巴巴”的地址为例，在仅使用文本召回的时候，几乎没办法找到相似的结果，如果通过对海量用户的点击行为进行分析，将点击行为加上地址文本信息合并形成高维向量，这样在检索的时候就可以天然的将点击率高的地址召回并排列在前面。

## 几乎覆盖了所有的 AI 场景

向量检索几乎覆盖了大部分的可以应用AI的业务场景。

例如广告反作弊、人群分析、视频指纹、版权保护、语音识别、智能问答、地址标准化、多语言翻译、地理信息、分子结构分析、生物群落分析等。



# 向量数据库的发展历程



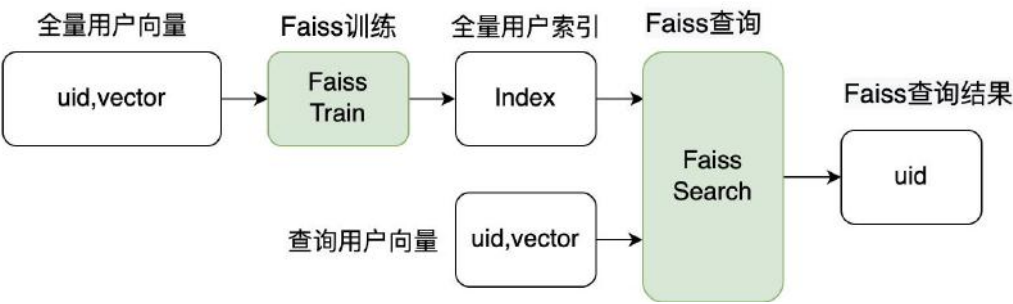
IDC 预测，到 2025 年，中国的数据量将增长到 48.6ZB，80% 是非结构化数据，并且将成为全球最大的数据圈。随着非结构化数据应用的增加，此类数据的处理分析需求也在随之增加。下面为向量数据库近年来的发展现状。



<p>2017年3月，Facebook 开源了 AI 相似性搜索工具 Faiss（Facebook AI Similarity Search）。支持相似度检索和聚类，多种索引方式，CPU 和 GPU 计算，以及 Python 和 C++ 调用。其使用场景最常见的为人脸比对，指纹比对，基因比对等。</p>	<p>2019年4月，Milvus 0.1 发布，2019年10月，Zilliz 开源了向量数据库 Milvus。Milvus 是一款开源的特征向量相似度搜索引擎。Milvus 使用方便、实用可靠、易于扩展、稳定高效和搜索迅速。</p>	<p>2019年10月，Vearch v0.1 发布。它是京东研发的一款分布式向量搜索系统，可以用来计算向量相似度或用于机器学习领域如：图像识别，视频识别或自然语言处理各个领域。Vearch 基于 Faiss 实现，提供了快速的向量检索功能。</p>	<p>Om-iBASE（向量数据库）是基于智能算法提取需存储内容的特征，转变成具有大小定义、特征描述、空间位置的多维数值进行向量化存储的数据库，使内容不仅可被存储，同时可被智能检索与分析。</p>	<p>TensorDB 是爱可生公司基于 Milvus 进行完善增强的企业发行版向量数据库软件。该产品实现了超大规模向量型数据的高效组织，设计了易扩展的索引结构，有效支撑了时变环境下的向量数据快速比对。</p>	<p>2021年10月，阿里巴巴发布了其开源项目一多模态向量检索引擎 Proxima。Proxima 是阿里巴巴达摩院自研的向量检索内核。目前，其核心能力广泛应用于阿里巴巴和蚂蚁集团内众多业务，如淘宝搜索和推荐、蚂蚁人脸支付、优酷视频搜索等。</p>	<p>2022年8月，Zilliz 推出了云端全托管向量数据库服务 Zilliz Cloud，进一步赋能企业 AI 应用，在全球范围内享有广阔的市场前景。</p>
--	--	---	--	---	---	---

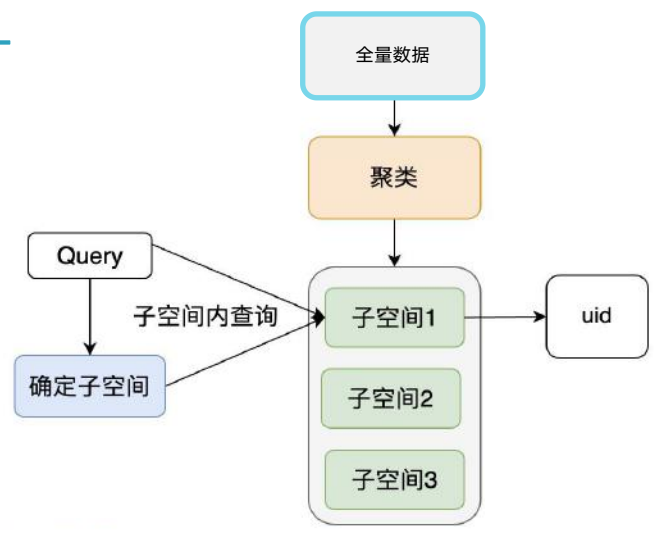
Faiss 全称(Facebook AI Similarity Search)是 Facebook AI 团队开源的针对聚类和相似性搜索库，为稠密向量提供高效相似度搜索和聚类，它包含一种在任意大小的向量集合中搜索直到可能不适合在 RAM 中的新算法。它还包含用于评估和参数调整的支持代码。Faiss 是用 C++ 编写的，带有 Python /numpy 的完整封装，并使用 GPU 来获得更高的内存带宽和计算吞吐量。

流程图



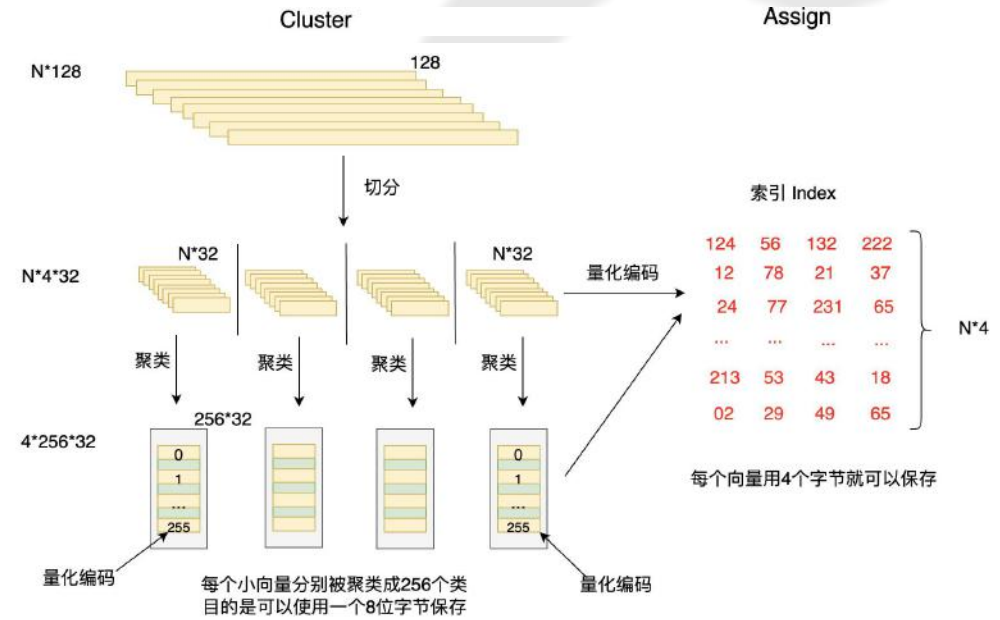
## 核心原理一

**倒排索引 IVF:** 通过计算 query 向量和所有子空间中心（如子空间内所有向量的均值）的距离，选出距离最近的 K 个子空间，表示和该 query 最相近的向量，最有可能在这几个子空间里。



## 核心原理二

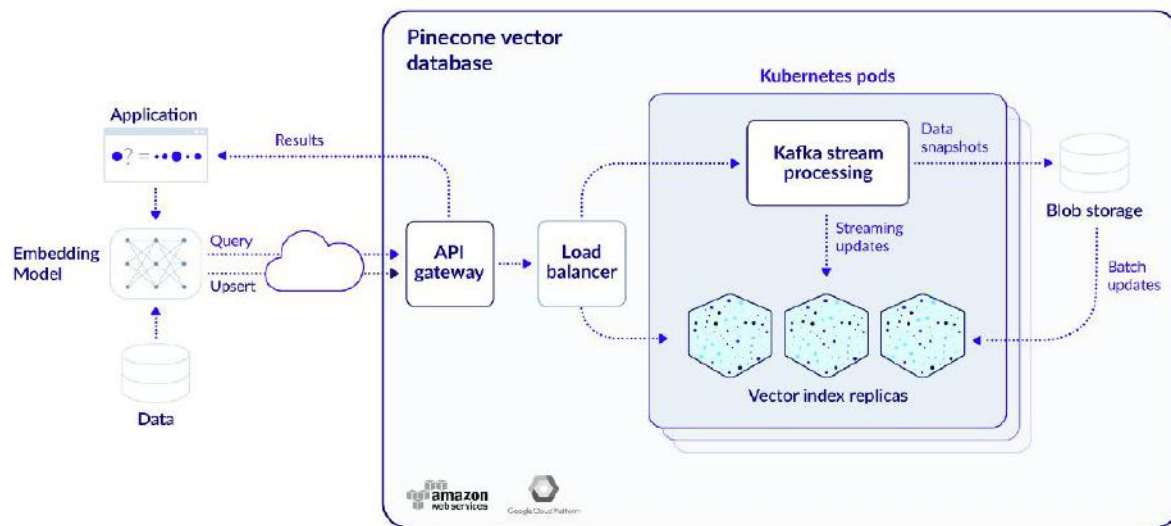
**乘积量化 PQ:** 乘积量化主要分为两个步骤，Cluster 和 Assign，也即聚类和量化。如图所示，假设每个向量的维度为128，每个向量被切分为4段，这样就得到了4个小的向量，对每段小向量分别进行聚类，聚类个数为256个，这就完成了Cluster。然后做Assign操作，先每个聚类进行编码，然后分别对每个向量来说，先切分成四段小的向量，对每一段小向量，分别计算其对应的最近的簇心，然后使用这个簇心的ID当做该向量的第一个量化编码，依次类推，每个向量都可以由4个ID进行编码。



# Pinecone - 商业全托管向量数据库

Pinecone 是一个托管向量数据库，它使开发人员可以轻松地将向量搜索功能添加到他们的应用程序中，只需使用一个 API。Gong、Clubhouse 和 Expel 等数百家公司使用 Pinecone 的向量搜索在其应用程序中添加了语义搜索、AI 推荐、图像搜索和 AI 威胁检测等功能。

Pinecone 架构图

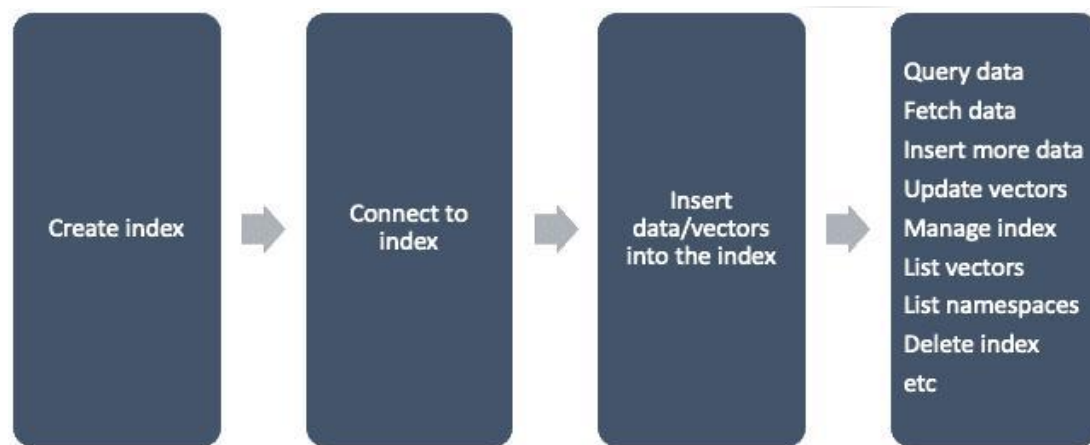


Pinecone 使用 Kafka 进行流处理，使用 Kubernetes 集群实现高可用性以及 Blob 存储（向量和元数据的真实来源，用于容错和高可用性）。

## Pinecone 特点

- **快速**：任何规模的超低查询延迟，即使是数十亿个项目。
- **实时**：添加、编辑或删除数据时实时索引更新。
- **过滤**：将向量搜索与元数据过滤器结合，以获得更相关、更快的结果。
- **完全托管**：易于启动、使用和扩展。

Pinecone 工作流程



**关键步骤**：创建索引 → 连接到索引 → 将数据（和向量）插入索引。

**功能**：查询数据（过滤数据）、获取数据、插入更多数据或更新现有向量、管理索引、管理数据。

## Pinecone 新功能

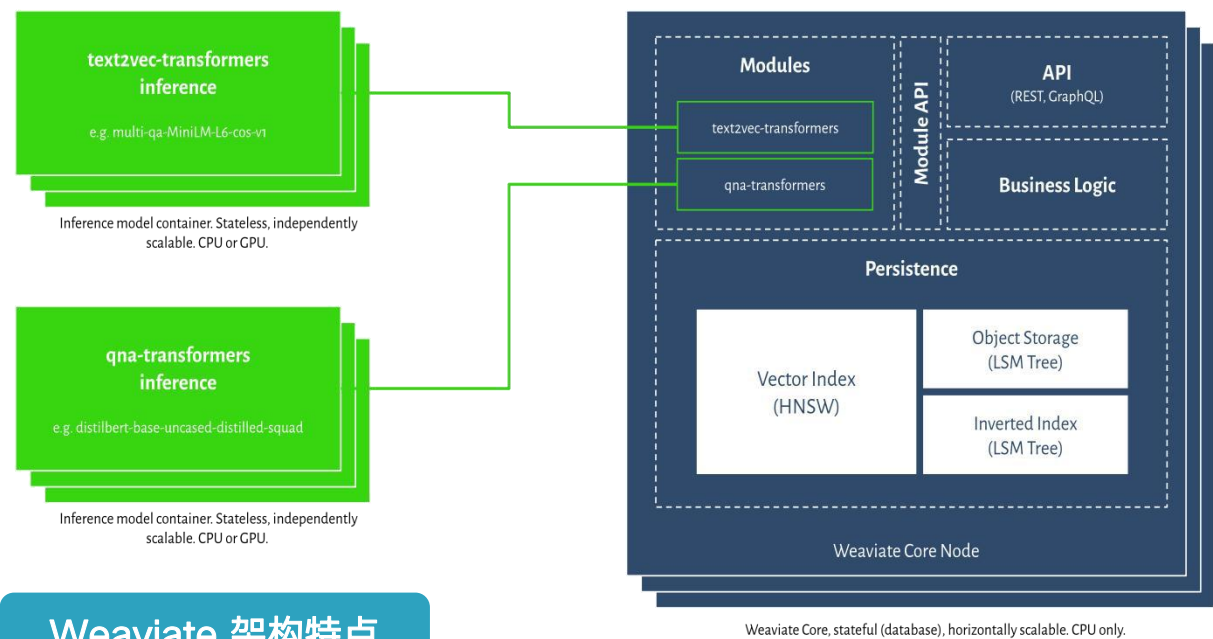
- **垂直扩展**：以零停机时间扩展您的向量数据块。
- **集合**：集中存储和重用向量嵌入和元数据，以试验不同的索引类型和大小。
- **p2 pod**：为高流量应用程序实现高达 10 倍的性能提升。

# Weaviate - 开源向量搜索引擎



Weaviate 是一个开源向量搜索引擎，它同时存储对象和向量，允许将向量搜索与结构化过滤相结合，并具有云原生数据库的容错性和可扩展性，所有这些都可以通过GraphQL、REST和各种语言客户端访问。该产品最重要的元素是向量搜索、对象存储和用于布尔关键字搜索的倒排索引的组合。

Weaviate 架构图



Weaviate 优势



• **低延迟和大规模**  
在任何规模的生产中实现低延迟。Weaviate 是一个容错、高可用性的数据库，您可以调整它以获得最佳的QPS/准确度权衡。



• **灵活性**  
将 Weaviate 用作独立的向量搜索引擎（带上您自己的向量）或使用众多模块之一（转换器、GPT-3等）来向量化或扩展您的 Weaviate 设置。



• **使用方便**  
云原生（Kubernetes 和 Docker）、各种客户端库，可通过RESTful和GraphQL API 获得。

## Weaviate 架构特点

### ● HNSW 性能提升

硬件加速和效率改进可将执行向量搜索或索引向量索引所需的时间减少多达 50%。

### ● LSM 树迁移

对象和倒排索引在 Weaviate 中的存储方式从基于B+Tree的方法迁移到LSM-Tree方法。这可以将导入时间加快 50%。

### ● 无需复制的水平可扩展性

由许多分片组成的索引可以分布在多个节点之间。一次搜索将触及多个节点上的多个分片并组合结果。

### ● 多分片索引

一个整体索引（每个类一个索引）可以分解成更小的独立分片。这允许更好地利用大型（单个）机器上的资源，并允许针对特定的大型案例调整存储设置。



# Proxima - 达摩院自研向量检索内核

Proxima 是阿里巴巴达摩院系统 AI 实验室自研的向量检索内核。目前，其核心能力广泛应用于阿里巴巴和蚂蚁集团内众多业务，如淘宝搜索和推荐、蚂蚁人脸支付、优酷视频搜索、阿里妈妈广告检索等。同时，Proxima 还深度集成在各类的大数据和数据库产品中，如阿里云 Hologres、搜索引擎 Elastic Search 和 ZSearch、离线引擎 MaxCompute (ODPS) 等，为其提供向量检索的能力。





# 全球向量数据库产业图谱

## 中国向量数据库产品提供商

 **zilliz**

 **JDT 京东科技**

 **蚂蚁金服**  
ANT FINANCIAL

( **Milvus** Manu)

( **VEARCH**)

(**ZSearch**)

 **LINKER 联汇**  
(Om-iBASE)

 **ACTION 爱可生**  
TECHNOLOGY  
(TensorDB)

## 国外向量数据库产品提供商

 **yahoo!**

( **vespa**  **Vald**)

 **{feature}form**  
(Embeddinghub)

 **drant**

 **Weaviate**

 **AquilaDB**


 **Pinecone**


 **TECHNOLOGY**


开源


商业


## 向量检索库


 **Meta**  
(Faiss)

 **Google**  
(ScaNN)


 **Microsoft**  
(SPTAG)


 **達摩院**  
ALIBABA DAMO ACADEMY  
(Proxima)


 **NSWLib**

 **Annoy**


## 向量插件


 **elastic**  
(elastiknn)


 **OpenSearch**  
(k-NN)


 **PostgreSQL**  
(imgsmir、CUBE、vops)

## 向量字段

 **Alibaba 阿里巴巴**  
(AnalyticDB-V)

 **蚂蚁金服**  
ANT FINANCIAL  
(Pase)

 **elastic**  
(dense\_vector)

 **Apache Solr**  
(DenseVector)

---

## 一、数据库排行榜及前沿动态

---

---

## 二、向量数据库的概述与解析

---

---

## 三、向量引擎加持传统数据库

---

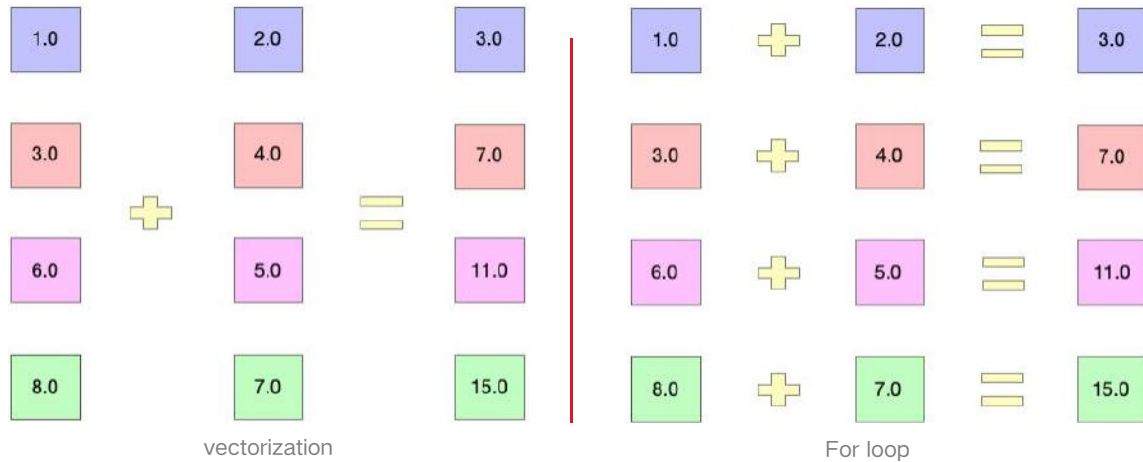
---

## 四、向量数据库产品相关案例

---

# 什么是向量化计算？

**向量化计算**(vectorization), 也叫vectorized operation, array programming, 指将多次for循环计算变成一次计算。向量化计算是一种特殊的并行计算的方式, 它可以在同一时间执行多次操作, 通常是对不同的数据执行同样的一个或一批指令, 或者说把指令应用于一个数组/向量。



左图中, 左侧为vectorization, 右侧为寻常的For loop计算。将多次for循环计算变成一次计算完全仰仗于CPU的SIMD指令集, SIMD指令可以在一条cpu指令上处理2、4、8或者更多份的数据。在Intel处理器上, 这个称之为SSE, 以及后来的AVX, 在Arm处理上, 这个称之为NEON。

因此简单来说, 向量化计算就是将一个loop——处理一个array的时候每次处理1个数据共处理N次, 转化为vectorization——处理一个array的时候每次同时处理8个数据共处理N/8次。

向量化在近年的OLAP列存数据库中大放异彩, 向量化引擎已经是目前主流OLAP数据库的标配, 它和Codegen一起决定了目前数据库执行层框架的架构设计和性能指标。向量化引擎的核心思想就是一次处理一批数据, 从而大大提高数据计算的速度, 例如对于一系列数据, 我们通过向量化技术可以一次处理1000行数据, 一次将这1000行数据做比较或者做加减运算, 这种处理方式在列存数据库上尤其有效, 因为列存数据库通常一列一列的将数据读取处理, 在内存中都是以Array的形式存储, 这种方式更容易使用向量化方式做计算。

## SIMD简介

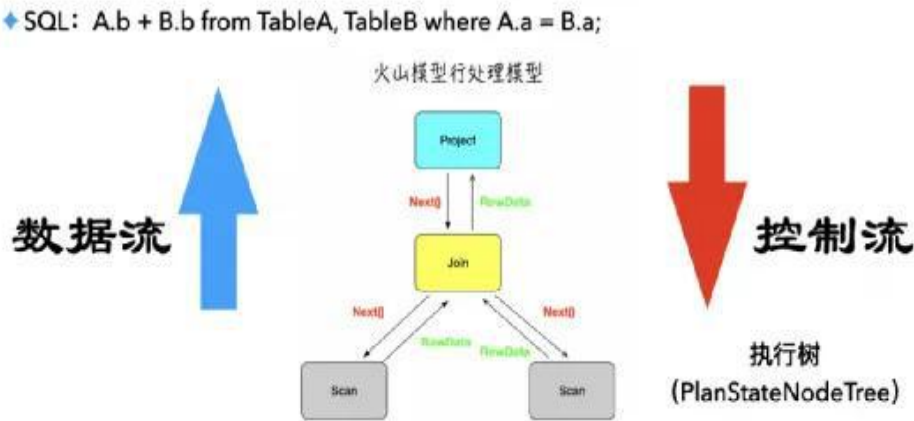
SIMD, 全称Single Instruction Multiple Data, 单指令多数据流, 能够复制多个操作数, 并把它们打包在大型寄存器的一组指令集。前提需要支持SIMD的CPU才能发挥其特性。单指令流多数据流, 也就是说一次运算指令可以执行多个数据流, 这样在很多时候可以提高程序的运算速度。以加法指令为例, 单指令单数据流(SISD)的CPU对加法指令译码后, 执行部件先访问内存, 取得第一个操作数; 之后再一次访问内存, 取得第二个操作数; 随后才能进行求和运算。而在SIMD型的CPU中, 指令译码后几个执行部件同时访问内存, 一次性获得所有操作数进行运算。这个特点使SIMD特别适合于多媒体应用, OLAP数据库等数据密集型运算。

# 为什么传统数据库需要向量化计算？



随着数据库软硬件技术的发展，经典的SQL计算引擎逐渐成为数据库系统的性能瓶颈，尤其是对于涉及到大量计算的OLAP场景。如何充分发挥底层硬件的能力，提升数据库系统的性能，成为近年来数据库领域的热门研究方向，而向量化执行就是解决上述问题的一种有效手段。

## 行式火山模型执行器



MPP数据库的API或者命令行接收到了SQL查询请求之后，系统先经过查询解析，然后进行查询优化，通过任务调度执行从存储引擎里面把数据读取出来，计算出结果集，返回给客户。一个查询语句经过词法分析、语法分析、语义检查后生成的结果叫做Query Tree，经过优化器之后的结果叫做Plan Tree。

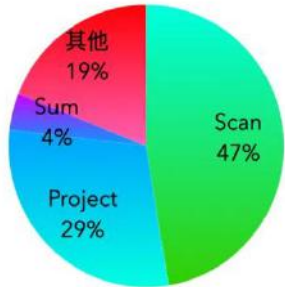
传统数据库执行查询计划通常采用火山模型的方式，流程如左图所示。火山模型具有简单、直观、易用等优点，早期数据库受限于硬件水平，IO、内存和CPU资源都非常昂贵，火山模型能够极大缩减内存使用量，因而被各大厂商普遍采用。如今，随着硬件技术的不断发展，火山模型的弊端也逐渐凸显。这种方式存在重复性执行多、反序列化代价高、数据局部性差等缺陷，而且一次执行仅处理一行数据，CPU花费大量时间在遍历查询操作树上，同时也没有针对CPU的SIMD能力等特性做优化，从而造成查询执行效率低下的问题。

向量化执行指的是将一次计算一条元组的形式，转换为一次计算多条元组的向量化计算。通过实现批量读取和处理，大大精简了函数调用开销，减少了重复运算，增加了数据的局部性，提高了执行效率。

- 向量化执行引擎可以减少节点间的调度，提高CPU的利用率
- 因为列存数据，同一列的数据放在一起，导致向量化执行引擎在执行的时候拥有了更多机会能够利用当前硬件与编译的优化特征
- 因为列存数据存储将同类型的类似数据放在一起使得压缩比能够达到更高，这样可以拉近一些磁盘IO能力与计算能力的差距

## PostgreSQL执行时间

SQL: Sum(a) from TableA;





火山模型的诞生为缓存数据库的内存压力，但该设计并未充分利用CPU的执行效率且以往的火山模型一次处理一个元组的方式造成过大的解释执行代价，阻止了对性能影响极大的编译优化。2005年《MonetDB/X100: Hyper-Pipelining Query Execution》的论文首次提出“向量化引擎”的概念，即为列存数据MonetDB设计一个新的执行引擎MonetDB/X100，使用向量化执行的方法，提高CPU使用率。

## 新的向量化执行引擎MonetDB/X100

**设计目标：**能够在执行大量的查询时达到较高的CPU使用率；可以扩展到其他应用领域，如数据挖掘和多媒体检索，并实现同样的高效率可扩展性代码；还能根据底层存储规模大小进行伸缩。

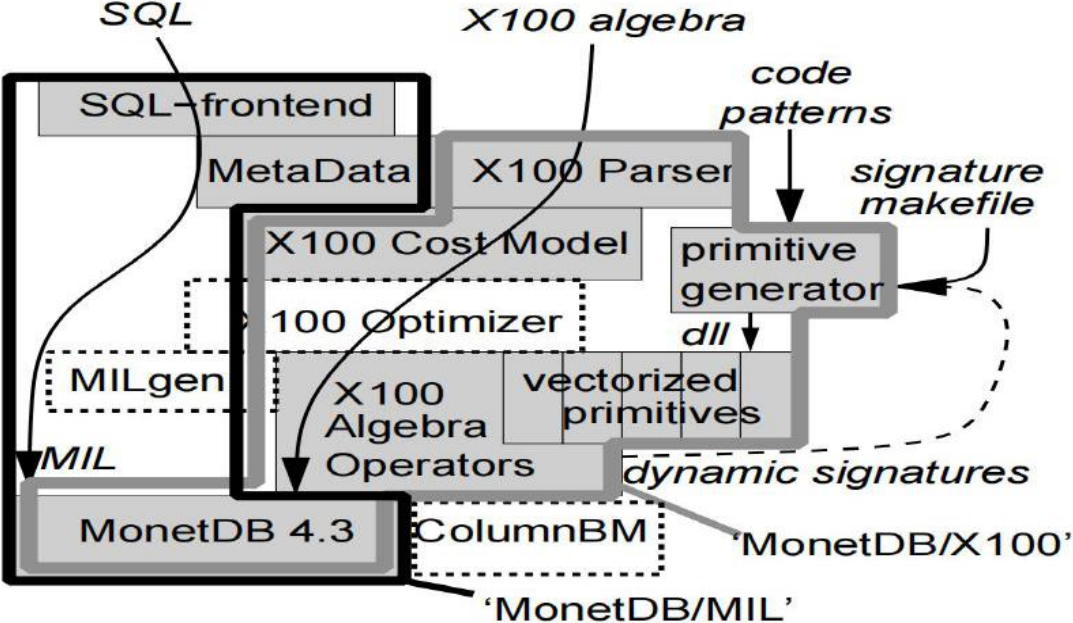
### X100架构设计克服了计算机架构中的瓶颈以提高性能

**Disk（磁盘）：**通过底层列存，减少不必要的数据存取，降低带宽要求，且可做基于列的轻量级压缩。

**RAM（内存）：**设计跟磁盘类似，也采取了列式存储的组织形式，目的也是为了减少内存占用和带宽压力。

**Cache（缓存）：**只有当数据进到缓存层才进行解压缩/压缩，这时由于cpu和缓存之间的高带宽,这种运算效率很高。X100的设计原则是尽量只在缓存块内做random数据访问。

**CPU（中央处理器）：**每个向量是针对一系列的切分一部分,向量化原语符合loop-pipelining的优化条件,可以重复利用CPU的并行流水线。而且可以针对复杂表达式,通过对向量化原语做组合,进一步提高执行效率。



X100架构设计



# 向量化执行引擎的技术价值

向量化执行引擎自 MonerDB-X100 (Vectorwise) 系统开始流行，现已成为在现代硬件条件下构建高效分析查询引擎。不同于传统模式，向量化实现了从一次对一个值进行运算，到一次对一组值进行运算的跨越。通过实现批量读取和处理，大大精简了函数调用开销，减少了重复运算，提高了执行效率。

01

## 减少虚函数调用，促进CPU乱序并发执行

传统查询执行引擎采用火山模型，按照一次处理一个元组的方式效率比较低。向量化查询执行引擎仍然采用火山模型，但是按照一次处理一组元组的方式，实现批量读取和批量处理，大大减少了函数调用开销，CPU可以把更多的时间集中到实际的计算上，效率会更高。

02

## 拉近一些磁盘IO能力与计算能力的差距

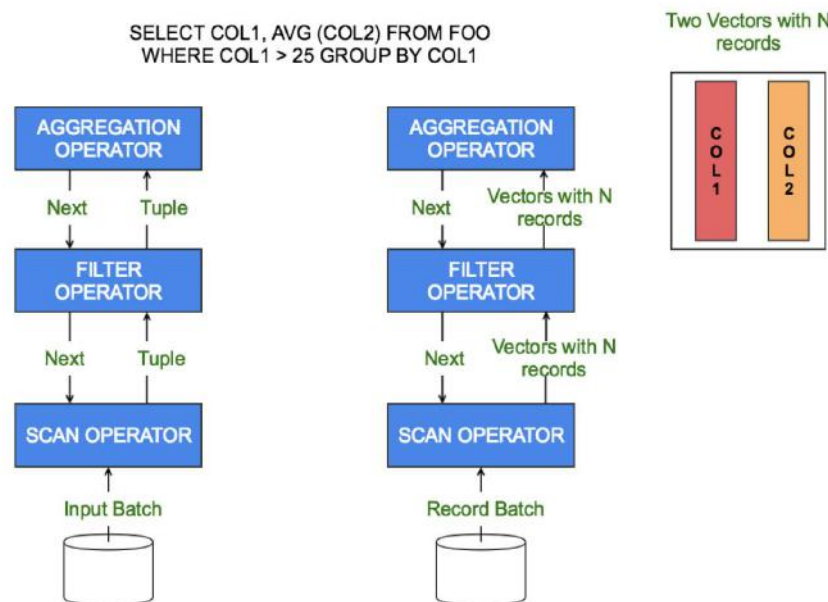
列存储技术在数据表的存储上使用数据表的列（记录的一个属性）为单位存储数据，这样类型一致的数据被放在一起，类似的数据在进行压缩的时候，能够达到一个比较好的压缩比。

03

## 更好利用当前硬件与编译的新优化特征

向量化引擎是跟列存储技术绑定的，因为列存储时每列数据存储在一起，可以认为这些数据是以数组的方式存储的。基于这样的特征，当该列数据需要进行某一同样操作，可通过一个循环来高效完成对这个数据块各个值的计算。相应的CPU的利用率得到了提高，另外数据被组织在一起，可以利用硬件发展带来的一些收益。

### 传统的一次元组处理 vs 向量化处理



# 向量化执行引擎设计实现和原理

## 设计实现

实现向量化的核心工作主要分为四块：

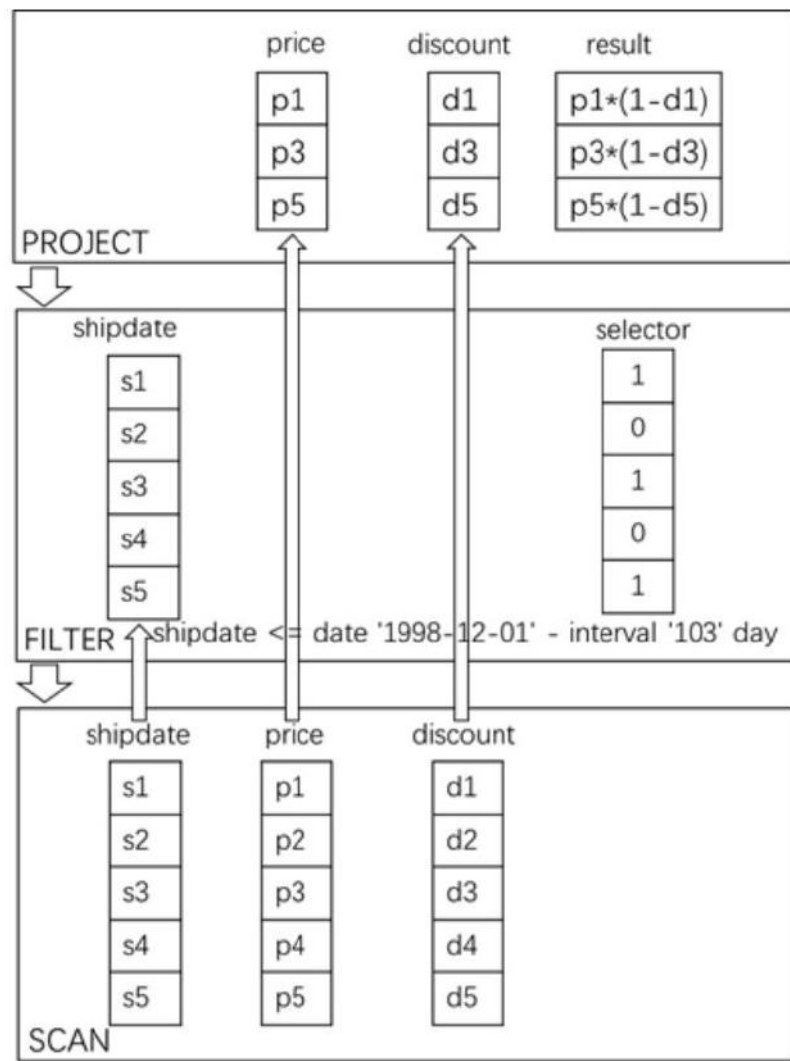
- **向量化执行框架**：为了让当前的执行器逻辑兼容向量化执行，需要考虑如何生成向量化计划，如何执行向量化计划，以及如何支持向量化执行和非向量化执行共存等。
- **向量化数据结构**：为了更好地发挥向量化执行的计算加速的作用，需要合理设计向量的内存组织形式。
- **向量化算子实现**：为了适应一次处理一组元组的执行方式，需要调整原有算子的实现。
- **向量化函数实现**：与向量化算子实现类似，向量化函数实现也要做相应的调整。

## 原理

向量化模型每次next调用返回的是一组元组，这样就可以将函数调用的代价均摊到多个元组上，从而减少总体函数调用次数。另外，算子内部实现或者计算函数实现可以使用更加高效的方式循环处理一组元组，比如使用编译器自动循环展开或者手动编写SIMD指令等方式。

需要注意的是，在实际的计算中往往执行的是在特定类型的列向量上的简单计算，连续的数据可以完全放入到cache中，计算过程中没有数据依赖以及条件分支，这样就可以充分发挥CPU乱序执行的能力，减少数据和指令的cache miss，从而将CPU硬件的能力充分释放。

## 向量化模型



向量化执行往往会带来更多的性能优势，因此在生成向量化计划时尽可能将计划路径中涉及的每个算子转换成向量化执行的方式。虽然没有通过比对所有计划代价并进行最优计划选择的方式，但是这种方式简单直接，而且可以通过GUC参数进行灵活控制。

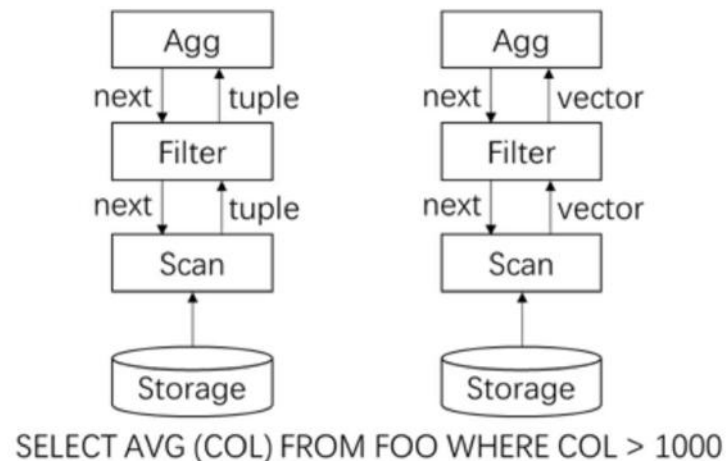
## 执行框架

一个查询计划生成之后，会尝试对其进行向量化转换。一般过程是先处理子计划，然后处理整个计划。对于每个计划节点，会根据计划节点的类型递归地对其包含的左右子树计划节点进行判断和转换操作，如果一个计划节点不支持向量化，可以通过在这个计划节点上面添加一个行转向量的新的计划节点，尽可能地让上层算子支持向量化执行。此外，还需要判断该计划节点包含的表达式计算是否支持向量化。比如对于Hashjoin计划节点，首先判断其左右子树计划节点是否支持向量化，如果都支持，则继续判断其包含的哈希键匹配函数以及非哈希键匹配函数等是否支持向量化，如果都支持，整个HashJoin就可以向量化执行。如果左子树计划节点不支持向量化，通过在其上添加一个行转向量的计划节点，使得HashJoin可以向量化执行。如果右子树计划节点不支持向量化，由于其对应Hash计划节点，与HashJoin计划节点是紧密关联的，故HashJoin节点不可以向量化执行，同时需要在支持向量化的左子树计划节点上面添加一个向量转行的新的计划节点，确保计划向量化计划和非向量化计划可以兼容。

**举例说明：**以一个简单的SQL为例，原有的火山模型执行流程和向量化模型执行流程如右图所示。两者都是上层算子迭代调用下层算子，但前者每次只能处理一个元组，而后者每次可以处理一组元组。

### 向量化执行要点：

- 采用vector-at-a-time的执行模式，即以向量（vector）为数据组织单位。
- 使用向量化原语（vectorization primitives）来作为向量化算子的基本单位，从而构建整个向量化执行系统。原语中避免产生分支预测。
- 使用code generation（代码生成）技术来解决静态类型带来的代码爆炸问题。

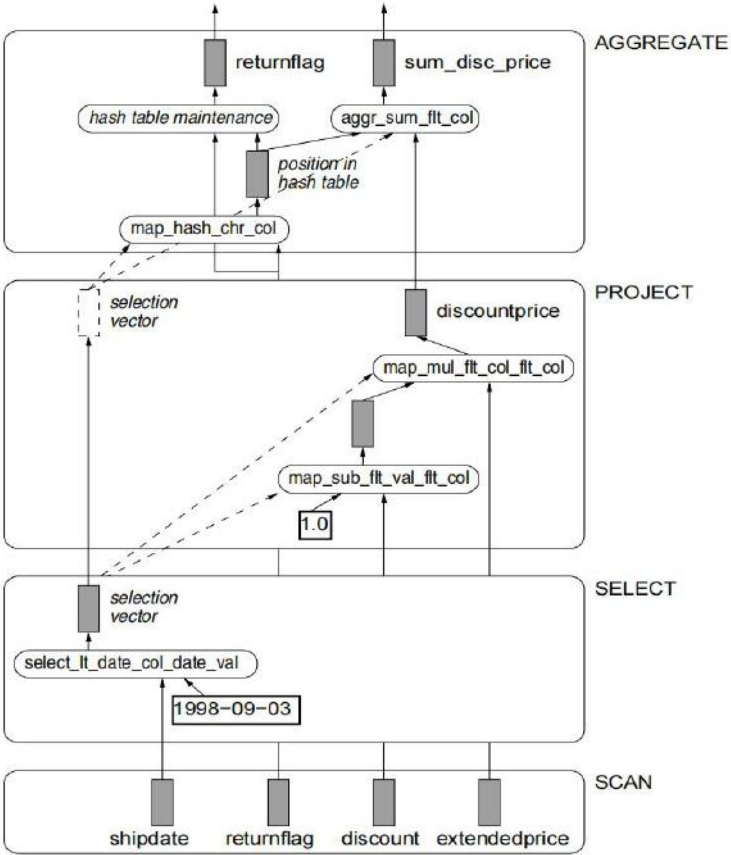


# 向量引擎MonetDB/X100 执行流程



MonetDB/X100的查询执行流程类似于原来的火山模型，主要区别在于执行粒度从原来的一个元组变成一个vector，函数调用次数大幅减少，同时对一个vector进行循环计算时可以用编译优化来提高CPU运行效率。另外，具体实现时需要增加一些辅助数组，以及对原有的执行逻辑的改造等。

MonetDB/X100 执行流程图



执行流程中包含了4个算子，即Scan、Select、Project和Aggregate：

- Aggregate计算主要包含两部分：计算每个元组在HashTable中的位置，计算聚集函数并将结果更新到对应的位置。新的位置需要在HashTable中创建。所有下层算子执行结束，不再生产新的vector后，遍历HashTable获取最终结果。Aggregate时也会使用selection-vector。
- Project主要是完成表达式计算并为最终聚集计算做准备。Project时会使用selection-vector跳过之前被筛选掉的元组，避免不必要的计算。
- Select创建一个selection-vector，在满足谓词条件的元组位置进行标记。这个数组在后面的计算过程中也同样会用到。引入selection-vector的好处在于，不必将筛选出来的数据进行拷贝和重新编排，允许在原来的输入向量上直接计算，效率会更高。
- Scan操作符每次从Monet BATs中检索数据向量。注意，实际上只扫描与查询相关的属性。

---

## 一、数据库排行榜及前沿动态

---

---

## 二、向量数据库的概述与解析

---

---

## 三、向量引擎加持传统数据库

---

---

## 四、向量数据库产品相关案例

---

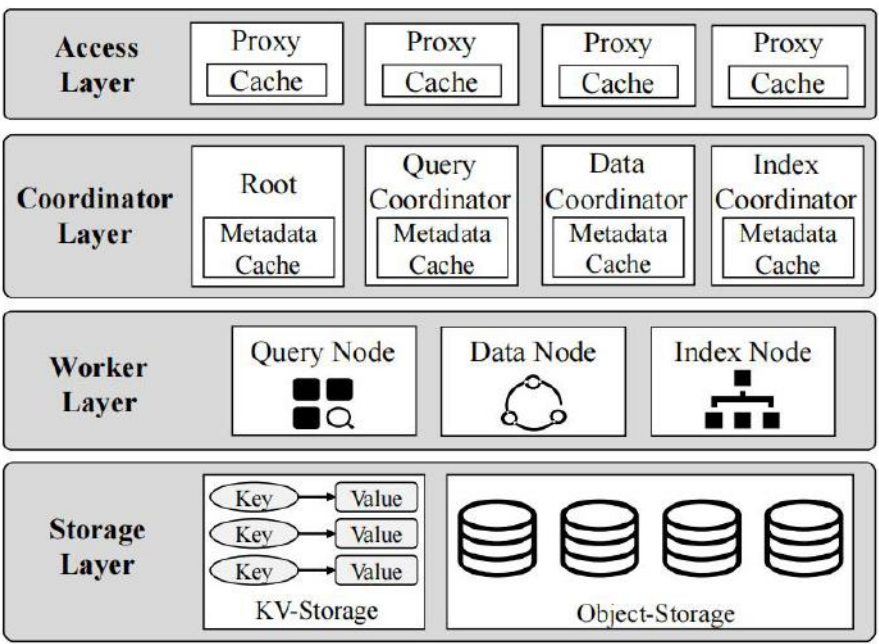


# Manu - 基于Milvus的云原生向量数据库平台



Manu 是 Milvus 2.0 的项目名称，它是由 Zilliz 创建的开源向量数据库。该向量数据库是专门为大规模处理向量数据而构建的，用于构建用于计算机视觉、NLP、定制搜索和新药发现等用例的人工智能应用程序。Manu 放宽了数据模型和一致性约束，以换取完全托管和水平可扩展的向量数据块所需的弹性和演变。Manu 还通过硬件感知实现和对复杂搜索语义的支持，对性能和可用性进行了广泛优化。

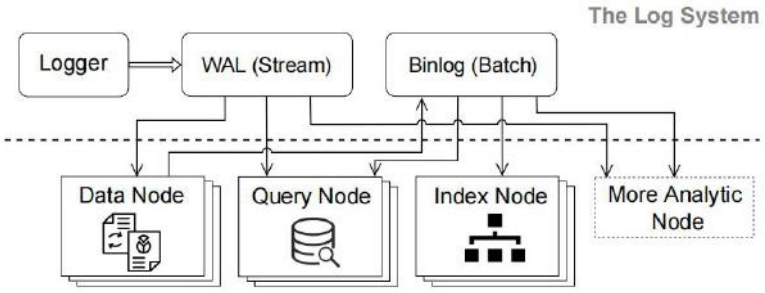
Manu 架构图



- **访问层：**由作为用户端点的无状态代理组成。它们并行地工作，以接收来自客户端的请求，将请求分发到相应的处理组件。
- **协调器层：**管理系统状态，维护集合的元数据，并协调处理任务的系统组件。
- **工作层：**执行实际的计算任务。工作节点是无状态的，它们获取数据的只读副本来执行任务，并且不需要相互协调。
- **存储层：**持久化系统状态、元数据、集合和相关索引。Manu使用 etcd（一个键值存储）来托管协调器的系统状态和元数据，因为etcd提供了故障恢复的高可用性。

Manu架构能够实现读取与写入、无状态与有状态以及存储与计算的解耦。

Manu 日志系统



- **日志系统：**Manu的核心枢纽，它连接着解耦的系统组件。Manu将预写日志（WAL）和 Binlog作为服务公开。
- 该日志系统是一个主要功能，允许每个组件独立扩展和演化，并促进资源分配和故障隔离。

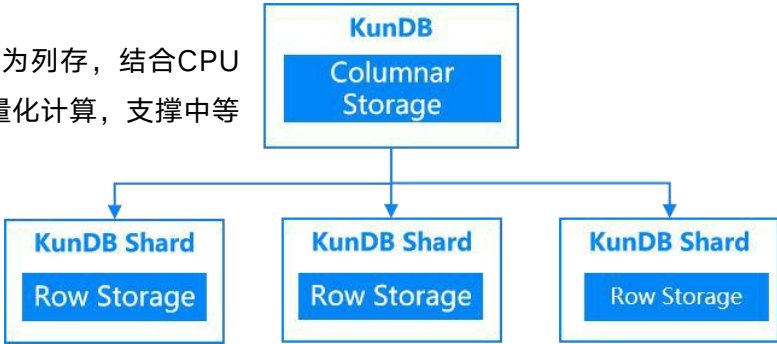
## Manu 采用 delta 一致性模型

该模型允许可调的一致性级别。增量一致性确保了更新的数据（包括插入和删除的数据）可以在 Manu 收到数据更新请求后，在增量时间单位内进行查询和搜索。

 Manu的5个目标：长期可进化性、可调一致性、良好的弹性、高可用性和高性能

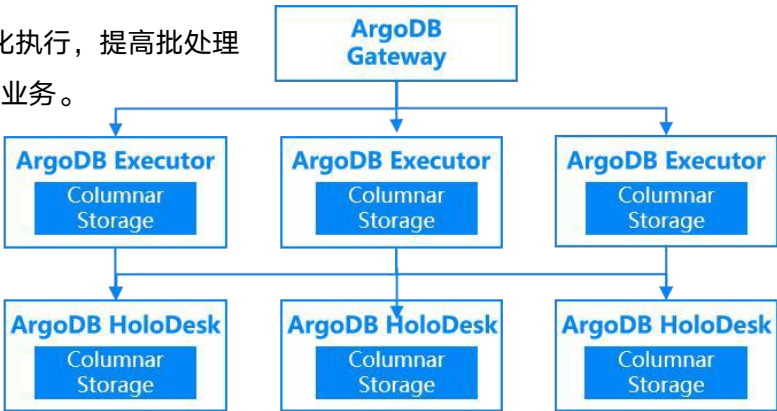
## ■ 分布式交易型数据库KunDB:

- 1.采用行存储提供高并发事务读写，保障高并发TP业务；
- 2.计算时在内存中转化为列存，结合CPU的SIMD指令分批做向量化计算，支撑中等数据规模的AP业务。



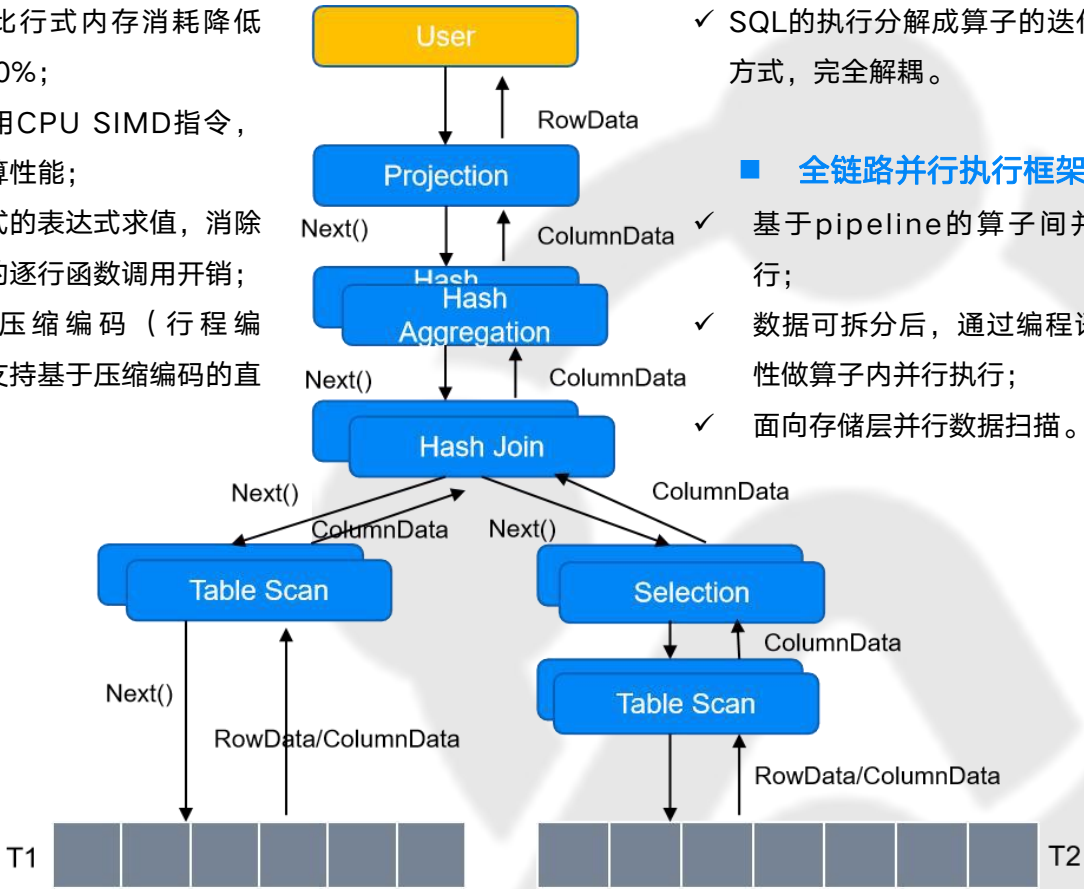
## ■ 分布式分析型数据库ArgoDB:

- 1.采用holodesk列存提供大数据量的高效存储；
- 2.分布式计算引擎向量化执行，提高批处理性能，支撑海量数据AP业务。



## ■ 向量化执行引擎:

- ✓ 内存列式存储，支持所有SQL数据类型；应用列式编码，相比行式内存消耗降低50%~80%；
- ✓ 充分利用CPU SIMD指令，提升计算性能；
- ✓ 面向列式的表达式求值，消除了行式的逐行函数调用开销；
- ✓ 自适应压缩编码（行程编码），支持基于压缩编码的直接计算。



## ■ 火山模型迭代执行:

- ✓ 常见的计算抽象成算子并专项优化，并持续扩展新型算子；
- ✓ SQL的执行分解成算子的迭代计算方式，完全解耦。

## ■ 全链路并行执行框架:

- ✓ 基于pipeline的算子间并行执行；
- ✓ 数据可拆分后，通过编程语言特性做算子内并行执行；
- ✓ 面向存储层并行数据扫描。

# Vearch - 基于 Faiss 的分布式向量搜索系统

Vearch 是京东研发的一款分布式向量搜索系统，基于 Faiss 实现，主要解决数亿级别向量的存储和计算查询的问题。可以用来计算向量相似度，或用于机器学习领域，如：图像识别、视频识别或自然语言处理等各个领域。它提供了快速的向量检索功能，能够提供类似 Elasticsearch 的 Restful API 可以方便地对数据及表结构进行管理查询等工作。

Vearch 整体架构分为以下3个模块：

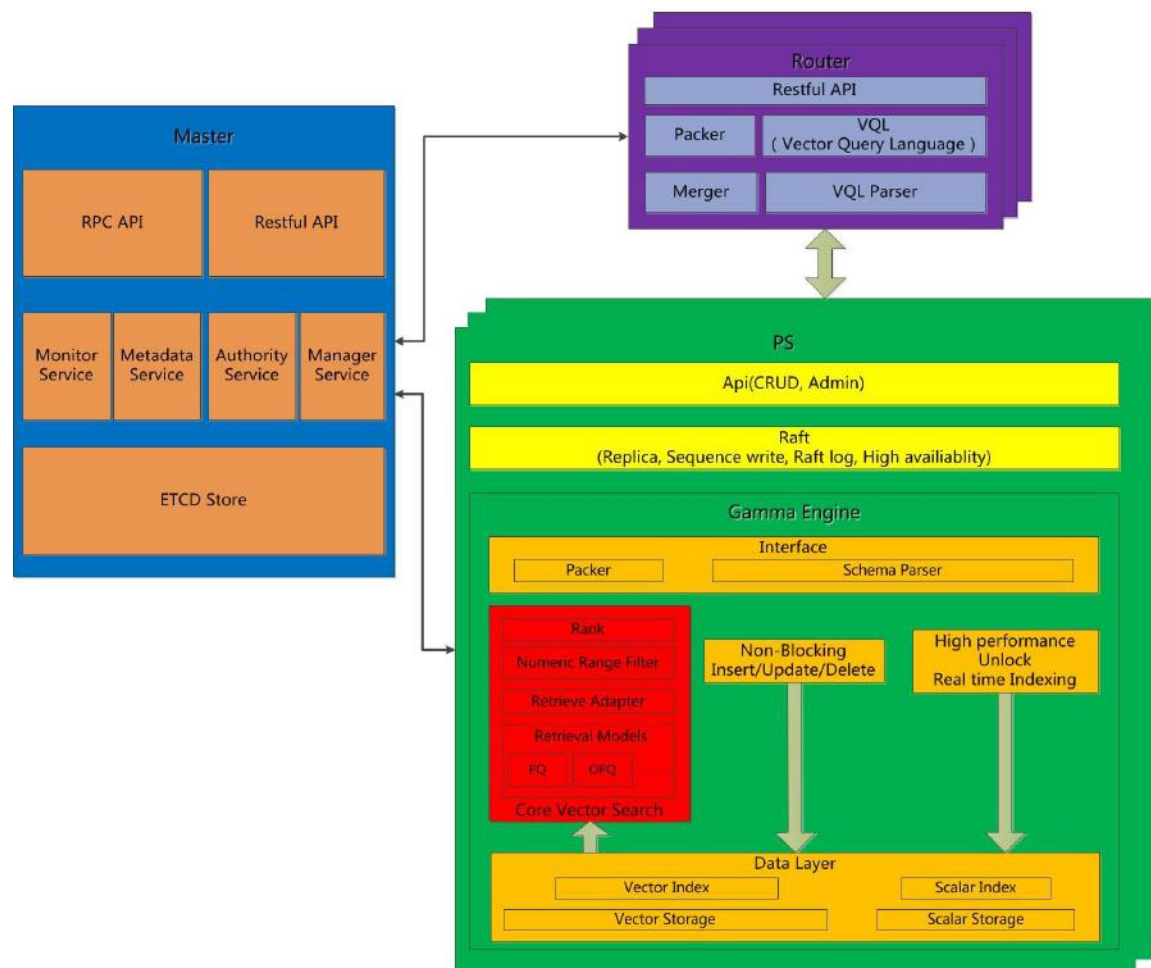
**Master**：负责元数据管理和集群资源协调。

**Router**：提供RESTful API: create、delete、search、update; 请求路由转发及结果合并。

**PartitionServer(PS)**：基于raft复制的文档分片; Gamma向量搜索引擎，它提供了存储、索引和检索向量、标量的能力。

## 系统功能

- 支持内存、磁盘两种数据存储方式；支持超大数据规模；
- 自研gamma引擎，提供高性能的向量检索；
- 基于raft 协议实现数据多副本存储；
- 支持内积(InnerProduct)与欧式距离(L2)方法计算向量距离。





# Om-iBASE - 基于智能算法的向量数据库

LINKER 联汇

墨天轮

Om-iBASE 是联汇科技开发的一款基于智能算法提取需存储内容的特征，转变成具有大小定义、特征描述、空间位置的多维数值进行向量化存储的向量数据库，使内容不仅可被存储，同时可被智能检索与分析。使用向量数据库可有效实现音频、视频、图片、文件等非结构化数据向量化存储，并通过向量检索、向量聚类、向量降维等技术，实现数据精准分析、精准检索。

## 向量化编辑器

支持对于文本、图片、视频、音频等多种多模态数据进行向量特征提取，并且包含丰富的语义信息，确保向量分析的全面性和检索的准确性。

## 高性能向量检索

通过自主研发的向量索引加速技术(ANN)，实现对于亿级别向量的秒级检索，有效支持高并发、大数据的向量应用场。

## 高性能向量分析

原生支持向量降维、数据聚类、异常分析等核心算法，通过GPU加速实现对于大量非结构化数据的实时分析。

## 灵活可配置且云原生

提供完整的SDK支持和灵活可配的插件体系，支持自定义向量化插件的热拔插接入，让开发者可以最大化发掘向量数据库的潜能。



# TensorDB - 基于 Milvus 的企业发行版

TensorDB 是上海爱可生信息技术股份有限公司基于 Milvus 开源向量数据库进行完善增强的企业发行版。该产品实现了超大规模向量型数据的高效组织，可以有效支撑时变环境下的向量数据快速比对，面向复杂场景下的实体分析与关系推断，克服了 AI 领域多样化应用面临的非结构化数据管理与处理分析困难，提升了数据库异构融合能力。并且具有极高的并发检索性能，支持卓越的水平拓展能力，能够满足多元业务场景下的高可用需求。

- 易扩展的索引结构：支持泛在时变场景下的向量型数据高效组织
- 高性能的并发查询：提供异构计算环境下的检索性能深度优化
- 低成本的混合查询：具备多类型混合字段的迅捷查询
- 高可用的系统架构：拥有金融安防等典型应用场景的引擎高可用解决方案



**内核引擎强大：**基于高性价比硬件，提供高性能的向量检索方案，数据插入与删除均在常数时间范围内，且服务器节点可以满足超大并发量下的高 QPS，满足用户复杂的应用场景。

**系统高可用性：**能够满足在业务不中断的情况下支持故障的自动切换与恢复，采用多副本的形式满足数据的可用性，保证用户业务系统平稳运行。

**系统按需扩缩容：**具备强大的负载均衡与水平扩缩容能力，对应用透明，系统拓展性强。

**索引实时更新：**支持大规模数据的实时索引构建，维护成本低，时变适应性优。

**混合字段检索：**引入灵活的字段拓展，实现了向量型数据与传统属性字段的混合检索，克服了传统基于异构数据库查询的效率瓶颈。

**异构计算支持：**利用 CPU/GPU 等异构计算平台资源优势，支持高吞吐量下的数据并发检索，最大程度地满足用户的低延时、高并发的需求。

互联网服务

工业制造

金融行业

安防行业

生物行业

地理服务

医疗行业

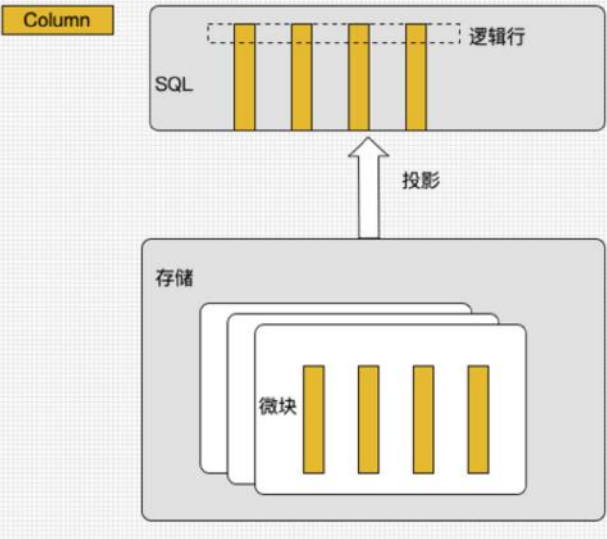
电子商务



为帮助客户解决 HTAP 混合负载下数据查询效率难的问题，OceanBase 引入向量化技术，并完全自主设计了向量化查询引擎，极大地提高了 CPU 单核处理性能，实现了 HTAP 场景下复杂分析查询性能的 10 倍提升。在 TPC-H 30TB 测试场景下，OceanBase 向量化引擎的性能是非向量化的 3 倍。OceanBase 向量化引擎的实现细节，主要包括存储和 SQL 两大方面。

## 存储的向量化实现

OceanBase 的存储系统的最小单元是微块，每个微块是一个默认 64KB（大小可调）的 IO 块。在每个微块内部，数据按照列存放。查询时，存储直接把微块上的数据按列批量投影到 SQL 引擎的内存上。由于数据紧密排列，有着较好的 cache 友好性，同时投影过程都可以使用 SIMD 指令进行加速。由于向量化引擎内部不再维护物理行的概念，和存储格式十分契合，数据处理也更加简单高效。整个存储的投影逻辑如下图：

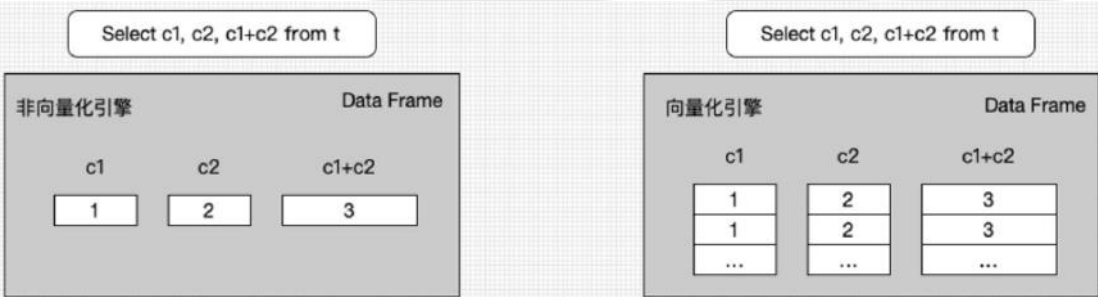


图：OceanBase 向量化存储引擎 VectorStore

## SQL 向量引擎的内存编排

SQL 引擎的向量化先前的数据组织和内存编排说起。在 SQL 引擎内部，所有数据都被存放在表达式上，表达式的内存通过 Data Frame 管理。Data Frame 是一块连续内存（大小不超过 2MB），负责存放参与 SQL 查询的所有表达式的数据。SQL 引擎从 Data Frame 上分配所需内存，内存编排如下图。

在非向量化引擎下，一个表达式一次只能处理一个数据（Cell）（下图左）。向量化引擎下，每个表达式不再存储一个 Cell 数据，而是存放一组 Cell 数据，Cell 数据紧密排列（下图右）




Doris实现向量化的核心工作，主要分为三块。首先是列式存储，需要在Doris本身的执行引擎当中，引入基于列存的内存格式。Doris存储层的本身就是以列方式存储的，但是在执行引擎当中，还是基于行的方式来做运算，是基于Tuple的运算，一次只能计算一个值，所以需要把对一个Tuple的计算切换为一个列式格式来计算，这样才能实现向量化。

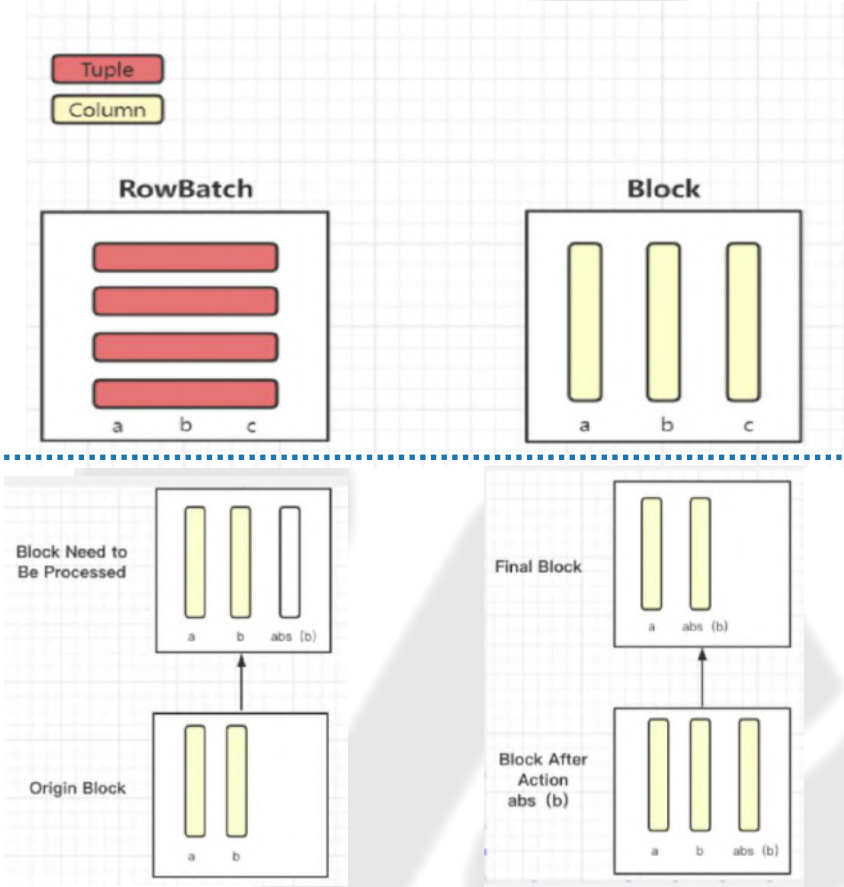
同时会基于新的列式存储格式，重新设计一套向量化和列式存储的计算引擎。

最后，基于列式存储跟向量化的函数计算框架，实现所有的向量化算子，包括当中常用的聚合、排序、JOIN，所有的SQL算子。

现有Doris执行引擎当中，内存结构如右图左边部分所示，是由一个RowBatch结构表示，它的数据是通过一个一个Tuple进行组织的，每一个Tuple是个连续的内存，可以看到左边RowBatch的结构，它分为三个列，但它每一行是连续的内存结构，每次也是处理一组值，但其实在这组值的内部处理当中，还是按行处理的。新的内存结构叫Block，它的数据是以列的方式来存储的，每一个列是一个连续的内存结构，现有Doris的执行引擎是RowBatch加Tuple实现的，新的向量化引擎变成Block和Column实现。


列式存储





在向量化函数执行框架当中，a列这部分内存是不再参与进来了，在内存结构上已经与b列独立开了，如右图所示，在b列经过abs计算之后，会生成一个abs(b)列，而a列在整个计算过程中都没有这个内存上的交互，做完abs计算之后，在原有的Block上新生成一个连续的内存结构，新生成一个列，是abs(b)列，那最后再把b列给过滤掉，最终的这个Block就留下a列和abs(b)列这两个列，就完成了一次列式存储的向量化计算。

计算框架



# TiDB - 向量化执行使表达式性能提升10倍



查询执行引擎对数据库系统性能非常重要。TiDB是一个开源兼容MySQL的HTAP数据库，部署广泛使用的火山模型来执行查询。当查询一个大库时，向量化模型会造成较高的解释开销以及较低的CPU CACHE命中率，向量化执行使用单指令在内存中执行一组连续的相似的数据项。与火山模型相比，向量化模型大大降低了解释开销。

## TiDB的SQL执行引擎完成了3项优化

01

在执行引擎中完成列式存储。类似Apache的Arrow。

02

将一次一个tuple的迭代模式（火山模型）改成一次一批。（1024个tuple）。

03

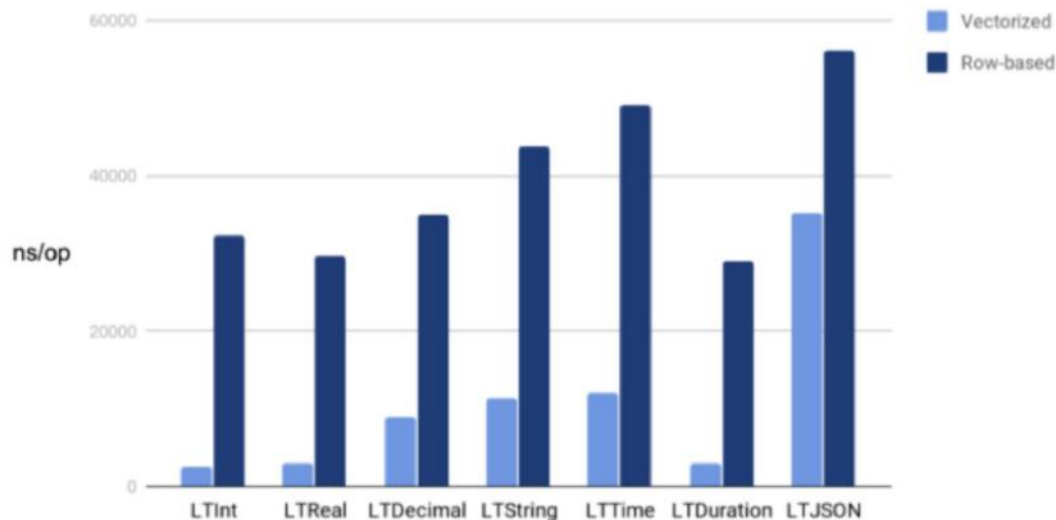
基于向量化执行原则，优化各种查询操作符。

## TiDB向量化执行基准测试

使用TiDB源码进行基准测试，并比较向量化前后的性能。使用相同数据（2列浮点数组成的1024行）分别 $col0 * 0.8 + col1$ 计算：

BenchmarkVec-12	152166	7056 ns/op	0 B/op	0 allocs/op
BenchmarkRow-12	28944	38461 ns/op	0 B/op	0 allocs/op

上面的结果表明向量化执行比基于行的执行引擎快4倍。下图对比了LT向量化前后各种小于（LT）函数的性能。横轴表示LT用于测试的函数，纵轴表示完成操作持续的时间（单位纳秒）



# MogDB - 向量化引擎加速OLAP系统



MogDB提供向量化引擎，通常用在OLAP数据仓库类系统。主要是因为分析型系统通常是数据处理密集型，基本上都是采用顺序方式来访问表中大部分的数据，然后再进行计算，最后将计算结果输出给用户。

## 向量化执行引擎简介

传统的数据库查询执行都是采用一次一数组（tuple）的pipeline执行模式，因此CPU的大部分处理时间不是用来处理数据，而是遍历查询操作树。这种情况下CPU的有效利用率不高，同时也会导致低指令缓存性能和频繁跳转。更加糟糕的是，这种方式的执行，不能够利用现代硬件的新优化特征来加速查询的执行。在执行引擎中，另外一个解决方案就是改变一次一数组（tuple）为一次一列的模式。这也是我们向量化执行引擎的一个基础。

向量化引擎是跟列存储技术绑定的，因为列存储时每列数据存储在一起，可以认为这些数据是以数组的方式存储的。基于这样的特征，当该列数据需要进行某一同样操作，可以通过一个循环来高效完成对这个数据块各个值的计算。

## 向量化执行引擎的应用

- **行存转向量化：**MogDB支持将行存表的查询转换为向量化执行计划执行，提升复杂查询的执行性能。通过对扫描算子增加一层RowToVec的操作，将行存表的数据在内存中变为向量化格式后，上层算子都能够转化为对应的向量化算子，从而使用向量化执行引擎计算。
- **使用向量化执行引擎进行调优：**为了提升行存表在分析类的复杂查询上的查询性能，MogDB数据库提供行存表使用向量化执行引擎的能力。通过设置GUC参数`try_vector_engine_strategy`，可以将包含行存表的查询语句转换为向量化执行计划执行。

## 向量化执行引擎的优势

- 可以减少节点间的调度，提高CPU的利用率。
- 因为相同类型的数据放在一起，可以更容易的利用硬件与编译的新优化特征。

## 向量化执行引擎的客户价值

通过批量计算，大幅提高复杂类查询性能。



# 往期报告免费下载



<https://www.modb.pro/doc/59620>



<https://www.modb.pro/doc/61120>



<https://www.modb.pro/doc/65548>



<https://www.modb.pro/doc/71694>



<https://www.modb.pro/doc/74438>



<https://www.modb.pro/doc/77118>