



Data Computing for New Discoveries

数据计算，只为新发现

如何从零开始参与技术社区

王刚，拓数派PieCloudDB 技术专家

大 纲

- 我与 OpenPie
- 社区贡献的从0到1
- 如何进行代码贡献

关于我

- 王刚 (Tender Wang)
- 拓数派 PieCloudDB 技术专家
- 多次参与PostgreSQL 代码贡献



拓数派：大模型数据计算系统先行者

- 拓数派（OpenPie）是立足于国内的基础数据计算领域**高科技创新机构**；
- 拥有强大的**数据库内核研发团队、数据科学团队和数字化转型团队**；
- 国内**虚拟数仓和eMPP技术提出者**，不断在数据计算引擎方向进行创新，**全面拥抱AI技术趋势**。

· 创虚拟数仓产品成为行业技术创新典范

拓数派专注数据计算领域

虚拟数仓打破

本报记者 李治国

“在大模型数据计算系统中，AI数学模型、数据和三者将前所未有地无缝链接、互为增强，创造出新的社会效益，成为助力社会高质量发展的新动力。”杭州拓数派科技有限公司（以下简称“拓数派”）创始人、董事长在接受记者采访时说。目前，拓数派已发布基于MPP分布式技术的存算一体机，适配信创，能够为企业降低运维开发时间；旗下虚拟数仓，可将物理数据迁移到云原生数据计算引擎，即将推出拓数派数据底座系统，成为AI技术底座，开启AI范式，助力企业充分发挥数据规模优势，赋能业务发展。

发明专利证书

证书号：第5775327号

发明名称：一种元数据访问方法、装置和存储介质

发明人：王源舟、杨峻峰、冯雷

专利号：ZL 2022 1 1418015.0

专利申请日：2022年11月14日

专利代理机构：杭州拓数派科技有限公司

地址：浙江省杭州市西湖区三墩镇三墩路133号（杭州拓数派科技有限公司）

邮编：311121

公开号：CN 11775327 A

公开日期：2023年7月14日

IPC分类号：G06F 16/28

发明人：王源舟、杨峻峰、冯雷

专利代理机构：杭州拓数派科技有限公司

地址：浙江省杭州市西湖区三墩镇三墩路133号（杭州拓数派科技有限公司）

邮编：311121

公开号：CN 11775327 A

公开日期：2023年7月14日

IPC分类号：G06F 16/28

数仓虚拟化

PieCloudDB 产品技术

数仓虚拟化由拓数派（杭州拓数派科技发展有限公司，又称“OpenPie”）于2023年3月正式提出。

数仓虚拟化^[1]可将物理数仓整合到云原生数据计算平台^[2]，根据数据授权动态创建虚拟数仓，按需灵活计算，打破数据孤岛，支撑更大模型所需的数据和计算^[5]。在云上，数据计算资源按需扩容，提升数仓的敏捷性和弹性，助力企业降低数仓管理复杂度^[3]，实现数量级增加可计算数据空间的同时，数量级降低数仓成本，打开无限数据计算空间，推进AI/BI到下一个精度^[4]。

中文名	数仓虚拟化 ^[1]	提出时间	2023年3月14日 ^[4]
技术突破：数仓虚拟化 ^{[1][2]}			
云原生存算分离架构 ^[3] ：运用元数据-计算-数据分离的三层架构，实现云上存储资源与计算资源的独立管理。云上计算资源弹性分配，有查询计算任务的时候按需启动，按照使用时间和规模计算成本。 ^{[1][1]}			
eMPP分布式专利技术^[3] ：在云上，PieCloudDB利用eMPP（elastic Massive Parallel Processing）架构，实现多集群并发执行任务。 ^{[1][1]} 企业可灵活进行扩容，随着负载的变化实现高效的伸缩，轻松应对PB级海量数据。			
全新的存储「精简」和缓存架构设计 ^[3] ：在计算层，各个计算节点针对元数据和用户数据都设计了多层缓存结构，避免网络延迟和数据移动，提高计算效率，保证用户的实时性需求。PieCloudDB针对底层对象存储设计了高效的文件格式，可在节省网络请求的同时提高计算效率。 ^{[1][1]}			
全新的优化器「达奇」 ^[3] ：PieCloudDB可以智能高效地生成统计信息，并生成更高效的查询计划。达奇优化器支持谓词下推，预计算，Block Skipping等高级特性，全面满足各种复杂的分析查询需求 ^{[1][1]} 。			

πDataCS 大模型数据计算系统技术突破

OpenPie®



πCloudDB

虚拟数仓服务HTAP | 点查



πCloudVector

向量数据计算



πCloudML

(大模型) 机器学习



...

大模型训练...



πFPGA 智能新硬件技术

数据存储 | 虚拟数仓 | 特定领域 (如神经网络)



Mundo元数据管理系统 统一Catalog



自研简墨存储

统一数据格式 | 一份数据多引擎计算 | 兼容主流云存储格式和协议

公有云

私有云

信创云

政务云

行业云

◆ 首创云原生eMPP架构

◆ 云时代存算双重分离技术

◆ 国内数仓虚拟化技术提出者

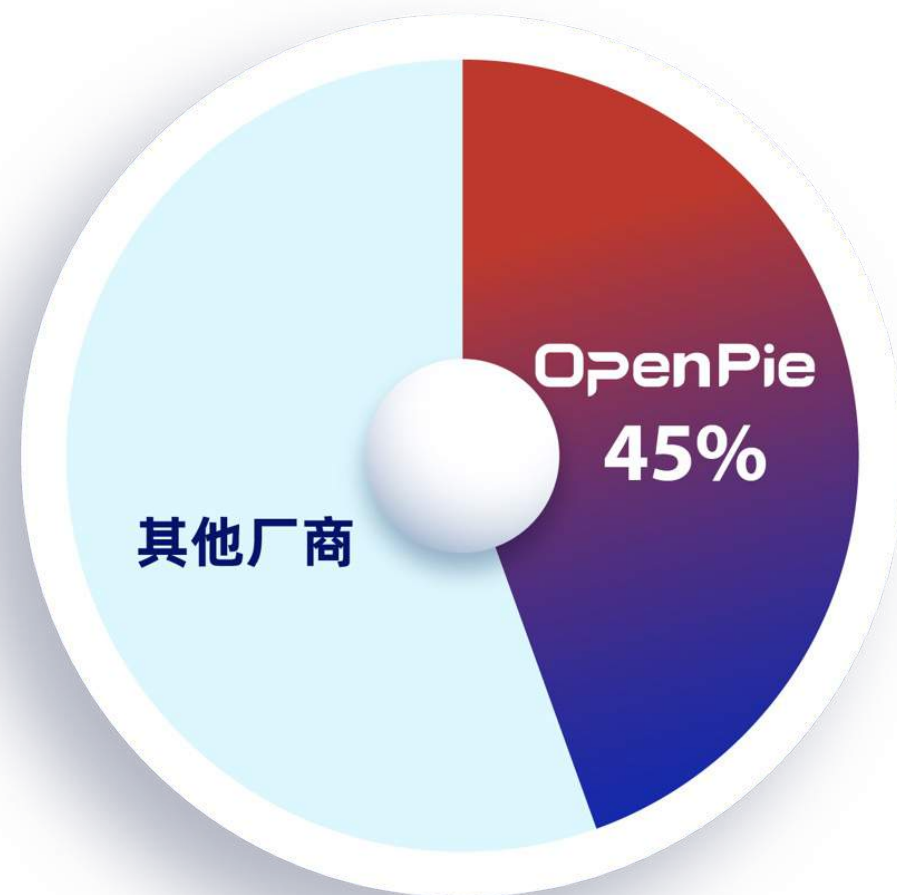
◆ 中国唯一全自研Table Format技术

核心产品优势

- 云原生存储架构，元数据、数据和计算全分离
- 数据一次入库永不出户，数据可用不可见，跑算力不跑数据
- 全链路加密保证数据安全
- DataSharing 技术原生支持数据要素流转
- TimeTraveling 等技术有效组织数据配合大模型精调
- 公有云、私有云及混合云，按需部署
- 国产软硬件一体机解决方案，满足特定行业用户需求

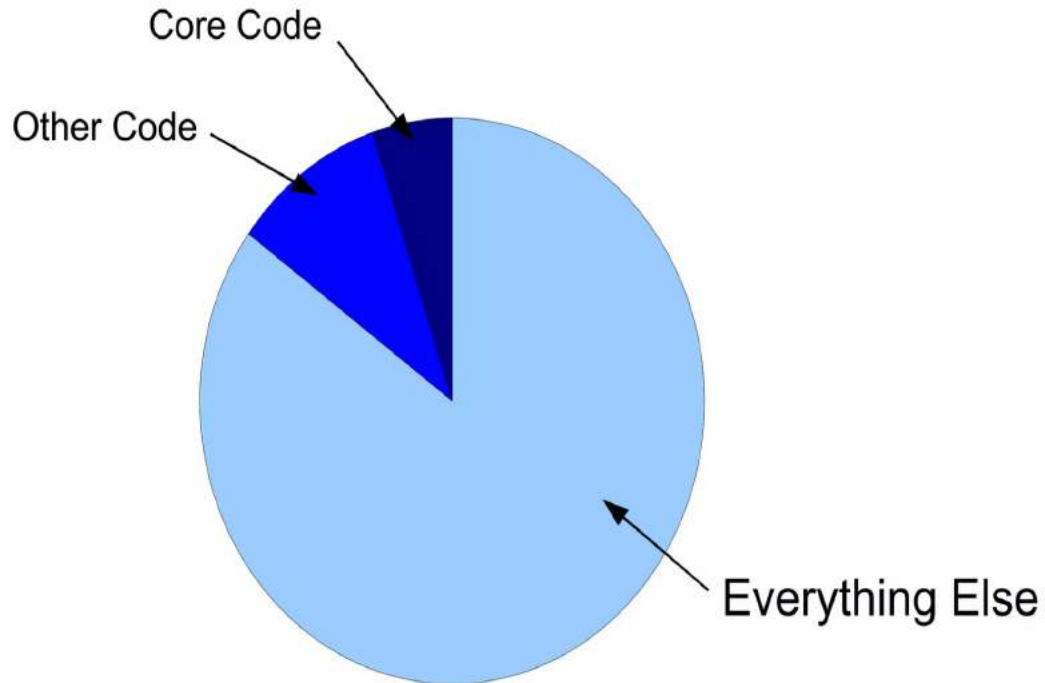
拓数派团队深耕国际技术社区影响力

国内总贡献次数	拓数派
278	124



How to Contribute to PostgreSQL

Contribution Pie



50 ways to love your project

**-- Josh Berkus
Core Team Emeritus**

How to Contribute to PostgreSQL

核心代码相关的贡献

- Bug report
- Review patch
- Write patch

How to Contribute to PostgreSQL

“周边”代码相关的贡献

- Contrib/
- Driver maintainer

How to Contribute to PostgreSQL

Non-code

- Translation
 - Main docs
 - Add-in docs
 - Website

PostgreSQL Need Writers

- Write main docs (pgsql-docs)
- Write add-in docs
- Write a HOWTO
 - Developer wiki, TechDocs
- Write a porting guide
- Write a case study
- Write a book
- Just blog it

How to Contribute to PostgreSQL

- User Groups

领导本地PG用户组，壮大PG社区

- Conferences

技术会议上做分享

Josh Berkus, 50 ways to contribute to PostgreSQL
<http://www.slideshare.net/PGExperts/50-ways-to-love-your-project>

How to Contribute to PostgreSQL

A Programmer's Perspective

On The PG Community

Mailing lists

- Everything is mail-based
- No bug tracker
- Hacker places
 - `pgsql-hackers`
 - `Pgsql-bugs`

How to Get Your Patch Accepted

Developing a patch for the POSTGRESQL project is a fairly complex process, and success is not guaranteed.

-- Bruce Momjian

Core Team Member

Implementation Details

- What subsystems and C files will be modified?
- How does it interact with other database features?
- How will error conditions be handled?
- Are there any system catalog changes?
- Are there any dump/restore issues? If so, fix pg_dump.
- Does psql need to be modified?

Sending a Patch

- Recommended: `git format-patch -1 -vN`.
 - Better when adding/removing files.
 - Better for large patch sets.
 - Includes commit message.
- Personal take: OK as long as `patch -pN` can work.
- [How to submit a patch by email](#)
- Discuss first, never rush writing code.

After Sending The Patch

- Everyone is watching

Coding in public can be uncomfortable.

- Be prepared to be humbled
- Sometimes you have to do it over
- Get agreement

Don't isolate yourself

Commit Fest

- Patch tracker:
 - review, commit or rejection.
 - Author one patch => Review one patch of “equal” difficulty.
- Includes bugs
- Needs community account.
- <https://commitfest.postgresql.org/>

Feedbacks You May Get

I don't like anything about the committed patch.

-- Tom Lane

Well, I don't think you should blame yourself, so no worries we are all here to learn :)

-- Michael Paquier

So Let's Begin to Hack The Postgres!

如何加入 PieCloudDB 社区

- 开源计划
- 参与方式
 - 产品试用（社区版、云上云版）
 - 贡献文章
 - 加入社区交流群
 - 等等



关注PieCloudDB公众号

随时获得产品动态



加入PieCloudDB技术群

获得更多技术干货

Thanks

Data Computing for New Discoveries

数 据 计 算 ， 只 为 新 发 现