

PieCloudDB: 基于PostgreSQL的eMPP 云原生数据库

吴疆 OpenPie产品和推广总监

吴疆

- OpenPie产品和推广总监
- 深耕云计算和数据库行业十余年
- 毕业于清华大学计算机系，先后在IBM, EMC, Pivotal, VMWare参与多个云平台 and 数据库项目



打造立足于国内 基础数据计算领域的世界级高科技创新驱动机构

OpenPie



杭州拓数派科技发展有限公司（又称“OpenPie”），以“Data Computing for New Discoveries”「数据计算，只为新发现」为使命，成立后的短短10个月时间内，完成了包括头部产业基金、东吴证券、元禾重元和政府科创平台在内的连续三轮战略融资。

旗下云原生分析型数据库PieCloudDB，以云计算架构为设计基础，首创全新eMPP分布式技术，帮助企业建立竞争壁垒的同时，实现数据价值最大化，并在新基建中承担可靠和可控的世界级云数据库底座。

Data Computing for New Discoveries
数据计算，只为新发现

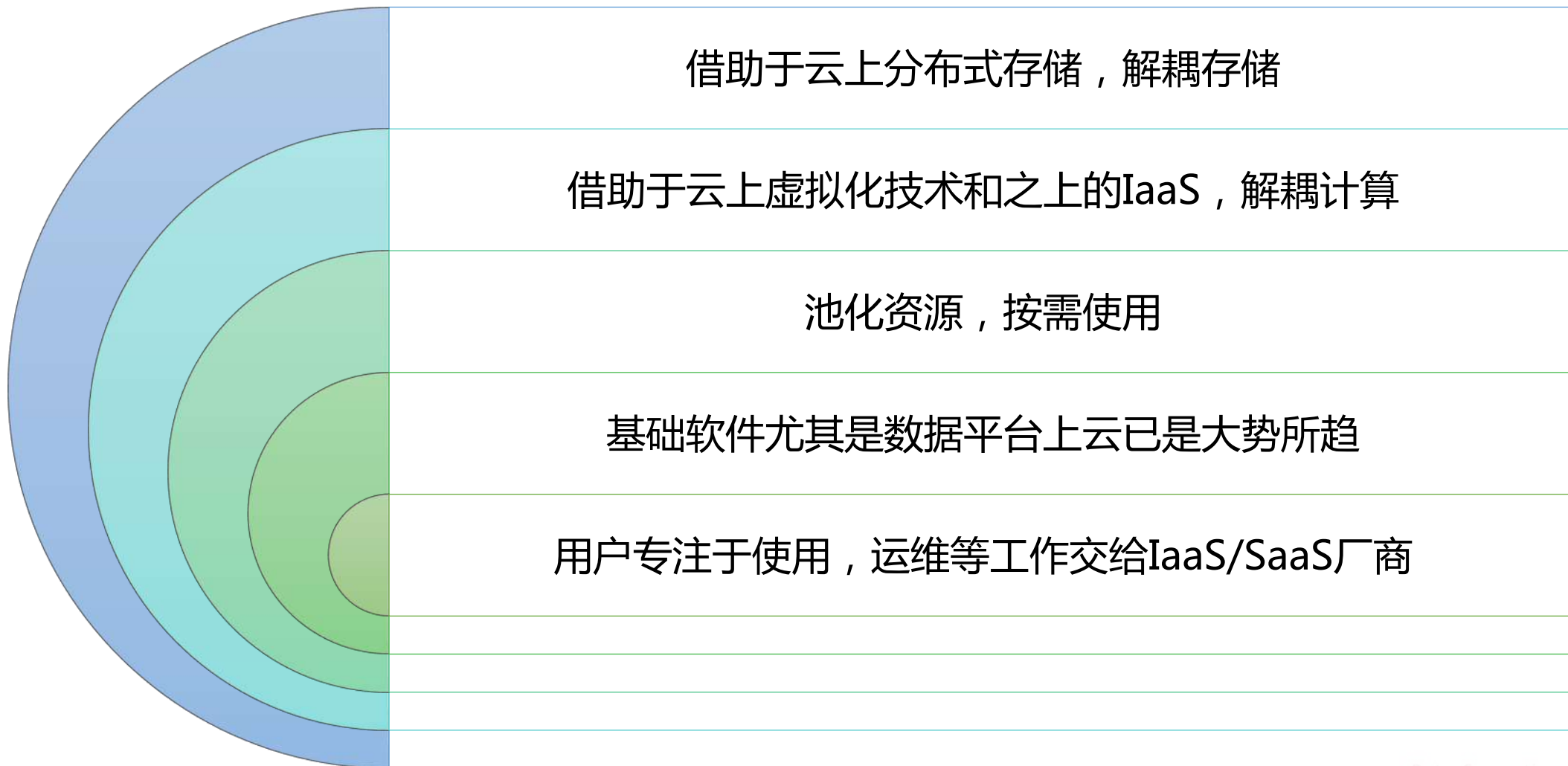
CONTENTS

- 1 数据库的云原生远景
- 2 云原生数据库PieCloudDB简介
- 3 PieCloudDB的架构特点
- 4 PieCloudDB 2.1版本新特性
- 5 总结

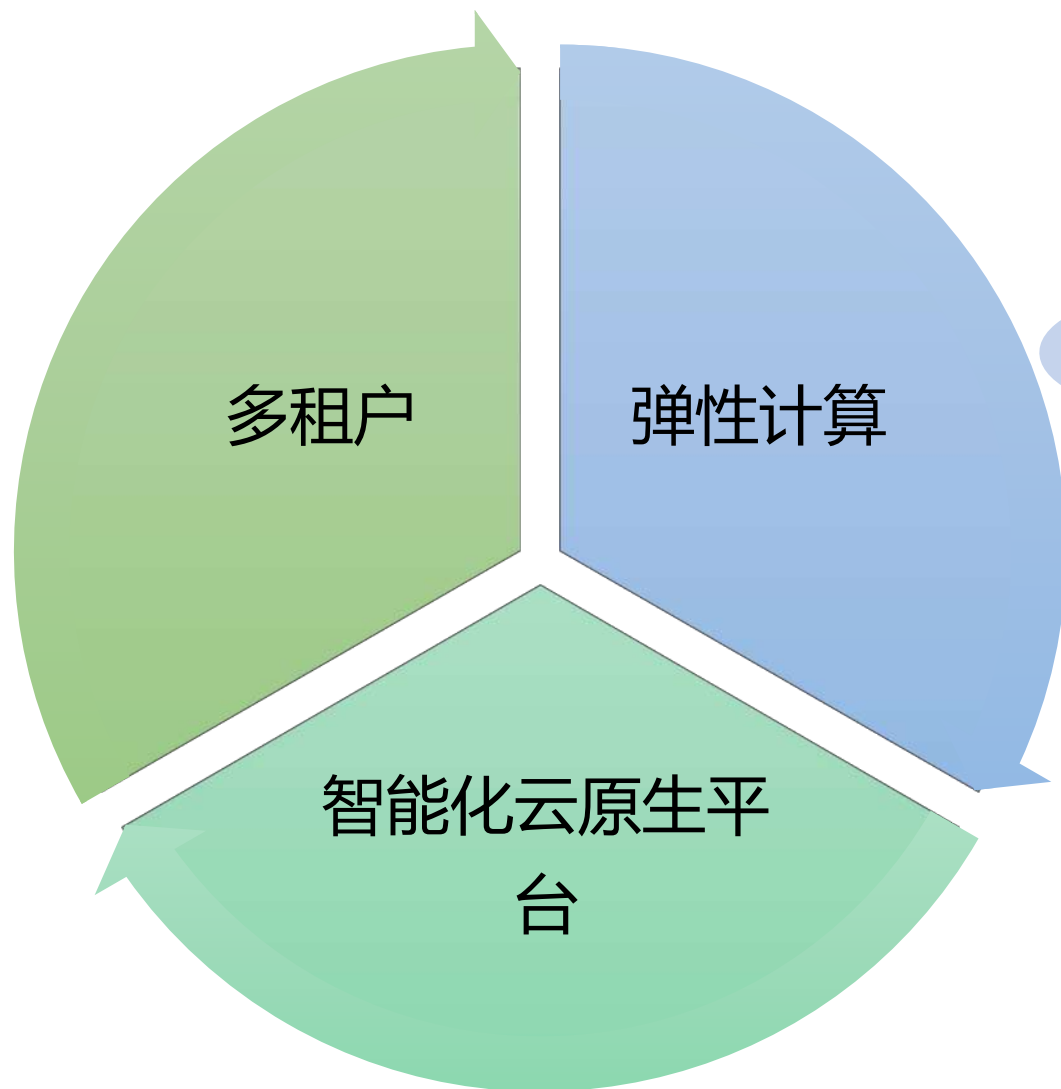
PART 01

数据库的云原生远景

云解决了什么？



上云 ≠ 云原生



- 产品要能支持存储资源和计算资源的分离
- 产品要能快速进行计算资源的弹性伸缩

传统分布式MPP架构痛点

缺乏弹性
业务使用不灵活

成本高昂
集群固定，资源利用率低

木桶效应
扩容难

数据孤岛
元数据和用户数据跨集群
访问困难

运维成本
运维和DBA



我们需要一个**云原生大数据平台**

PART 02

云原生数据库PieCloudDB简介

关于PieCloudDB

OpenPie

一个云原生实时大数据平台

平台底层：eMPP 云原生分布式SQL数据库

我们的目标：支持多模，serverless的实时大数据平台

● 安全可靠

● 使用简单

● 功能齐全

● 性能极致

Data Computing for New Discoveries
数据计算，只为新发现

PieCloudDB 重要特点



eMPP

- 弹性计算资源（横向和纵向）、极速调整
- 共享用户数据（典型如廉价对象存储）
- 共享元数据
- MPP架构：分布式，海量数据并行处理



友好的用户接口（WebSql, ODBC/JDBC driver等）.



云原生



完备的事务支持



完善的SQL标准支持



安全



Postgres生态支持

Postgres 生态

PieCloudDB 重新打造 PostgreSQL 12.x 实现存算分离

PieCloudDB 对几乎所有内核模块做了大量的创新

PieCloudDB 内核团队拥有强悍的Postgres内核代码掌控能力

团队也拥有丰富的Postgres内核大版本升级合并经验

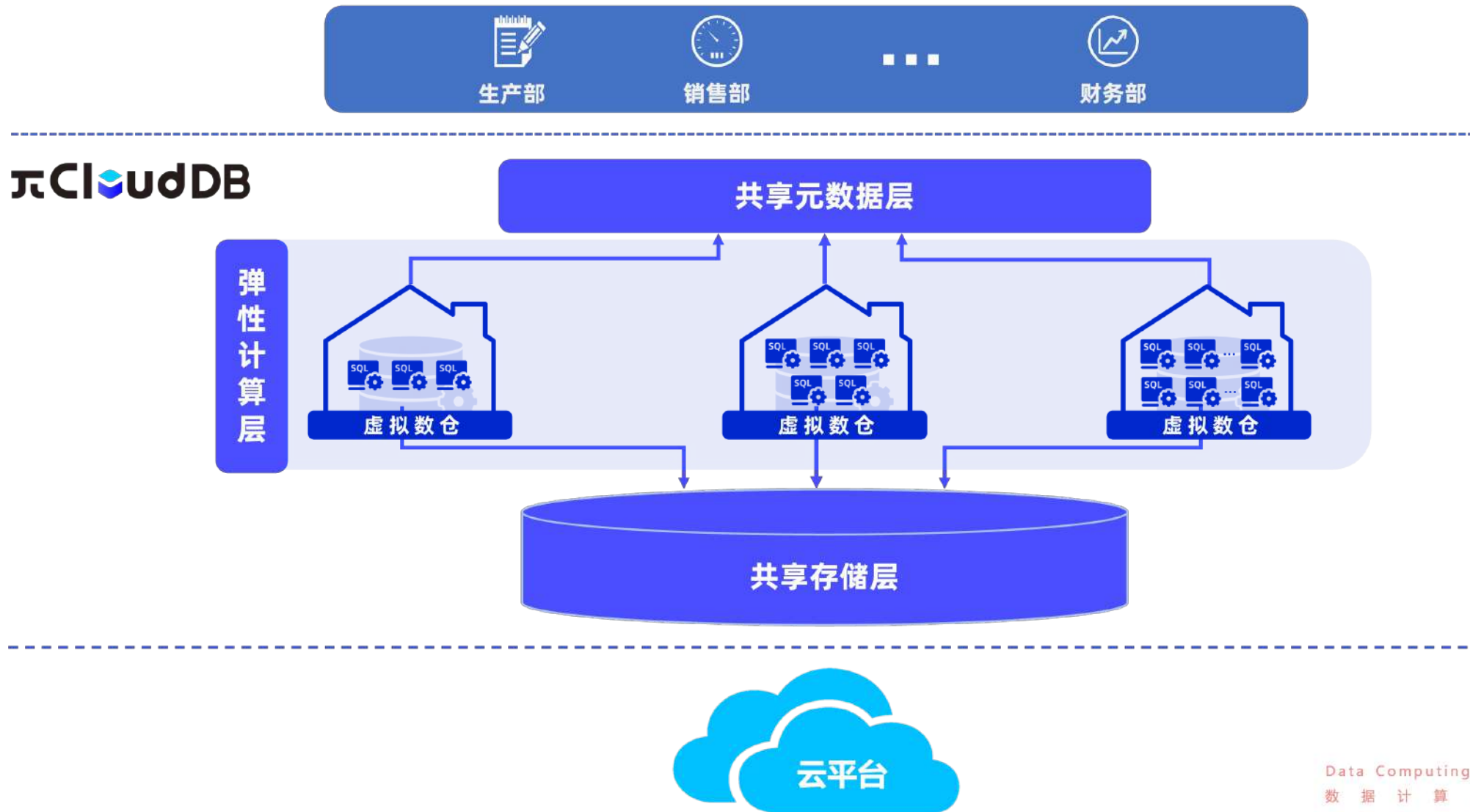
- 将来会保持和Postgres内核大版本对齐

PART 03

PieCloudDB架构特点

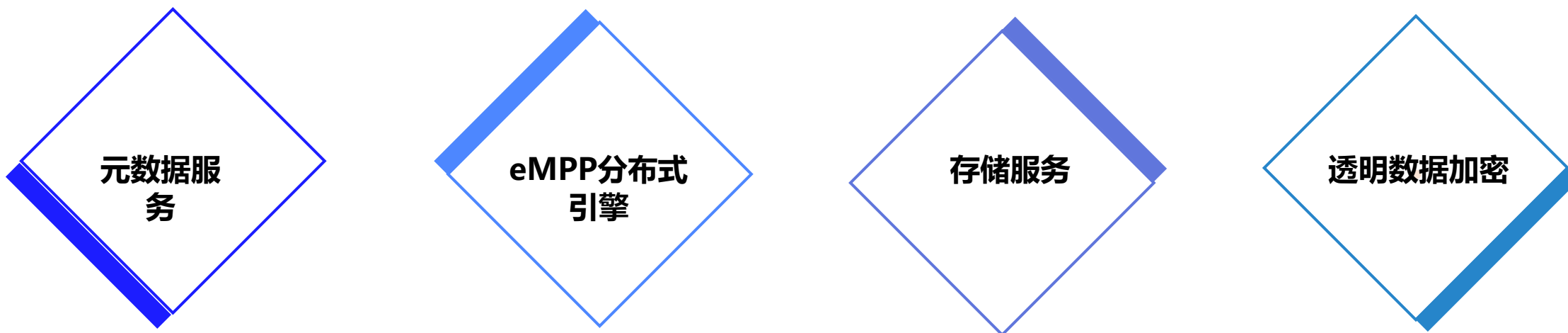
PieCloudDB 架构

OpenPie



PieCloudDB 核心架构特点

OpenPie



Data Computing for New Discoveries
数据计算，只为新发现

01 | 元数据管理

元数据管理的设计目标

- 高可用和多集群
- Multi-master

实现多节点共同访问的数据存储

实现分布式锁

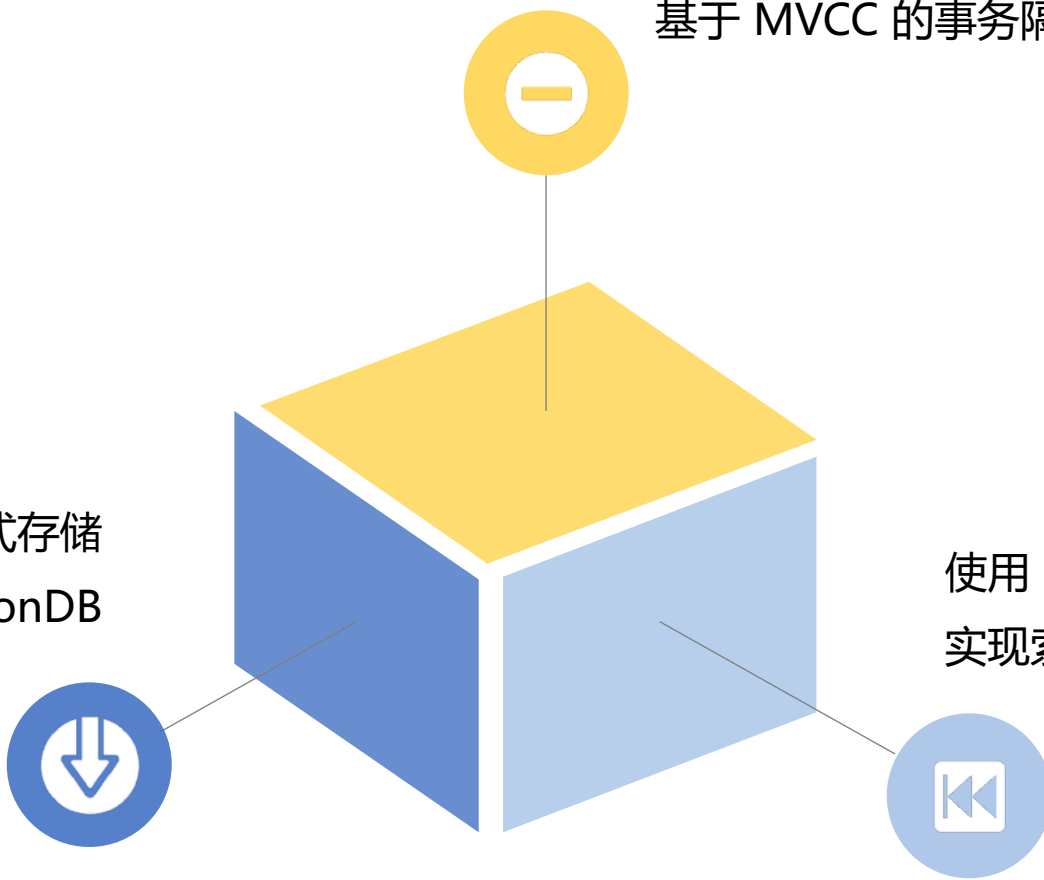
- 多机并发访问
- 分布式环境下的多版本

mstore — FoundationDB上的Catalog

基于 MVCC 的事务隔离级别

将元组以 key-value 的形式存储
到 FoundationDB

使用 FoundationDB Key 的自然排序
实现索引



mstore — FoundationDB上的Catalog

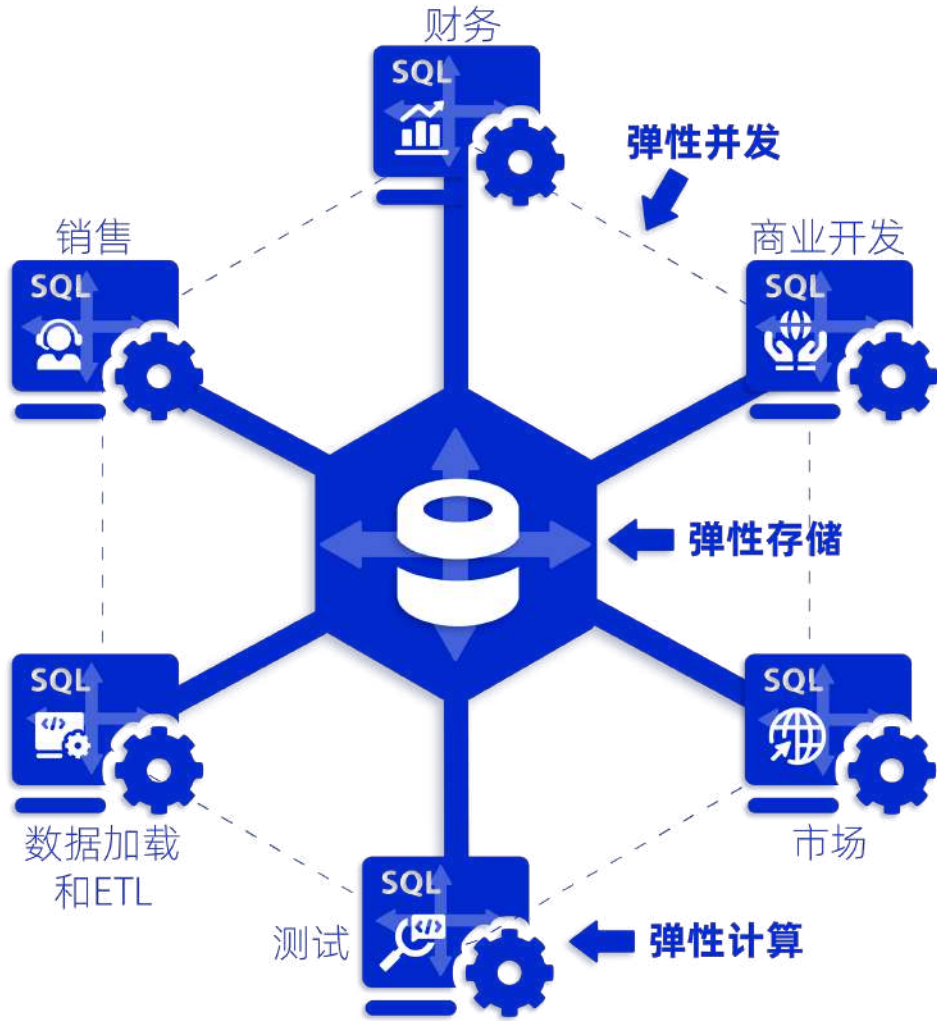
使用和 Postgres 相同方式存储元数据 —— 将元数据存储到系统表中

实现新的基于key-value的存储来存放系统表

02 | 分布式引擎

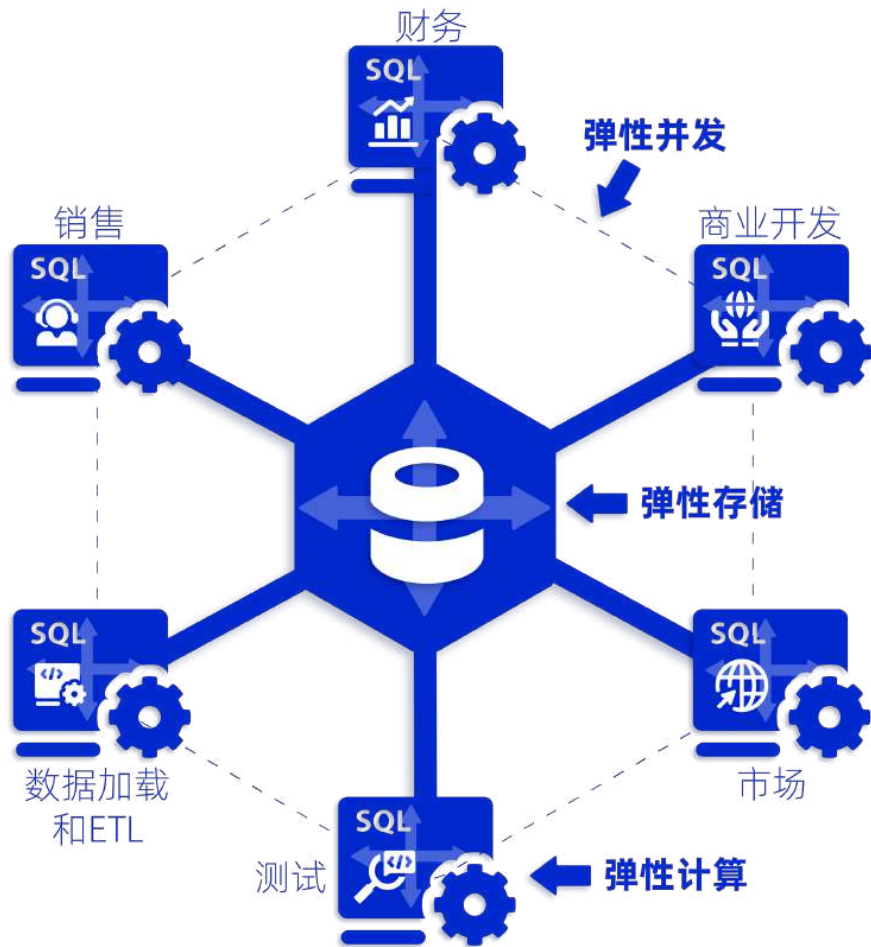
计算

- MPP
 - 将一个单一计算任务在大量独立的计算机上并行执行。
- 多租户、多集群
- 弹性伸缩：集群大小、集群类型、集群数量
- 隔离性：不同租户、不同负载
- 高并发
- 高可用
- 可按使用量付费



存储

- 多租户隔离
- 容量和带宽独立于计算伸缩
- 可按使用量付费
- 高可用/可靠存储
 - 支持跨多数据中心复制数据
- 唯一真理
 - 全局只需要存储一份数据，通过共享存储来实现数据共享，避免拷贝和维护多份数据副本



事务

- ACID
 - 支持两种隔离级别：读已提交、可重复读
- 扩展性
 - 事务管理器无单点性能瓶颈
- 隔离性
 - 不同租户之间的事务管理器是完全隔离的，不会相互影响
- 容错性
 - 事务管理器支持对各类基础设施故障进行自动容错

03 | 用户数据存储

构建新一代云原生存储引擎

- Multi-Cloud 云上设施
 - 对象存储（数据共享，存算分离）
 - 兼容HDFS，NAS，本地磁盘
 - 公有云，私有云，混合云
- 现代硬件
 - CPU/GPU 高速缓存访问
 - 数据的局部性优化（SIMD）
 - 现代存储技术
 - 新硬件的使用



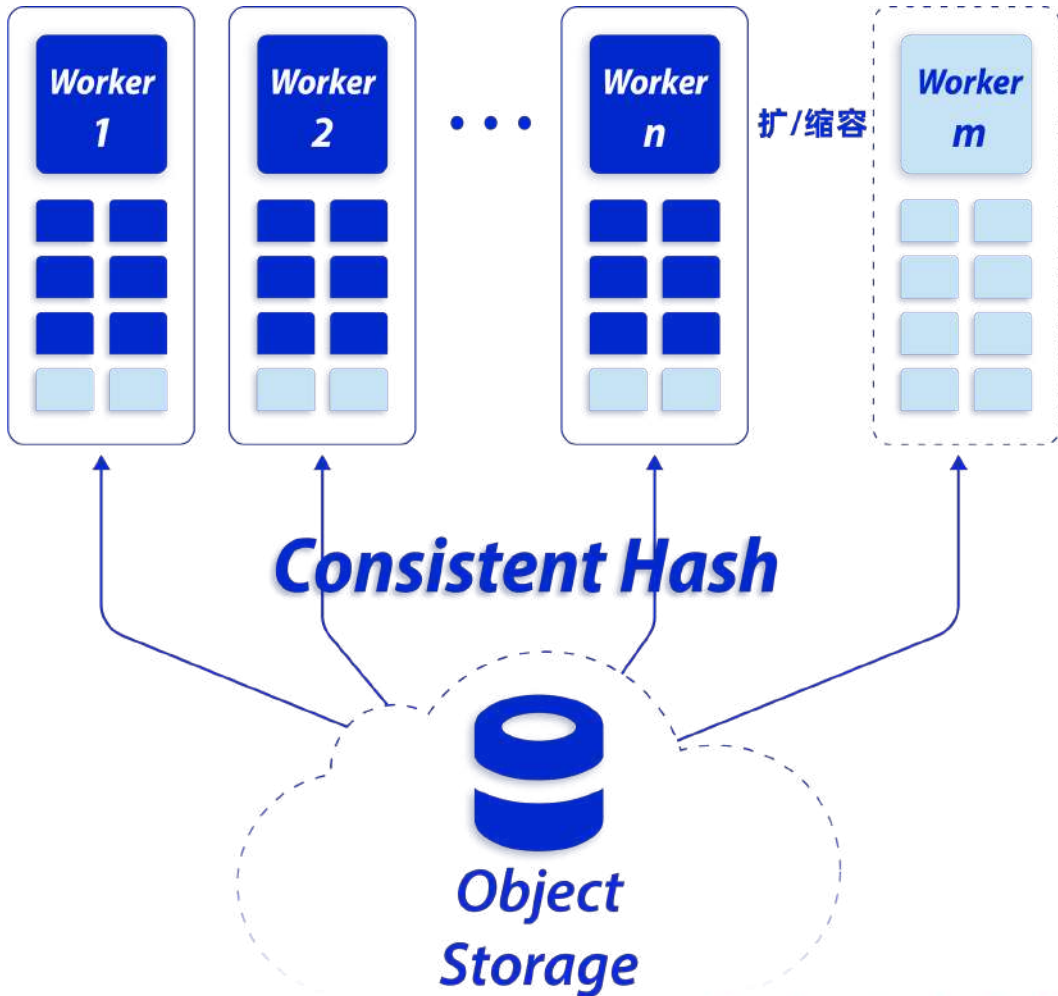
构建新一代云原生存储引擎

- 数据分布和弹性

- 分布式eMPP架构 (一致性Hash)
- 本地数据减少高延时的云存储访问
- 减少数据移动
- 扩缩容最少的数据移动

- 数据安全性

- 透明数据加密
- 三级密钥
- 实时加解密



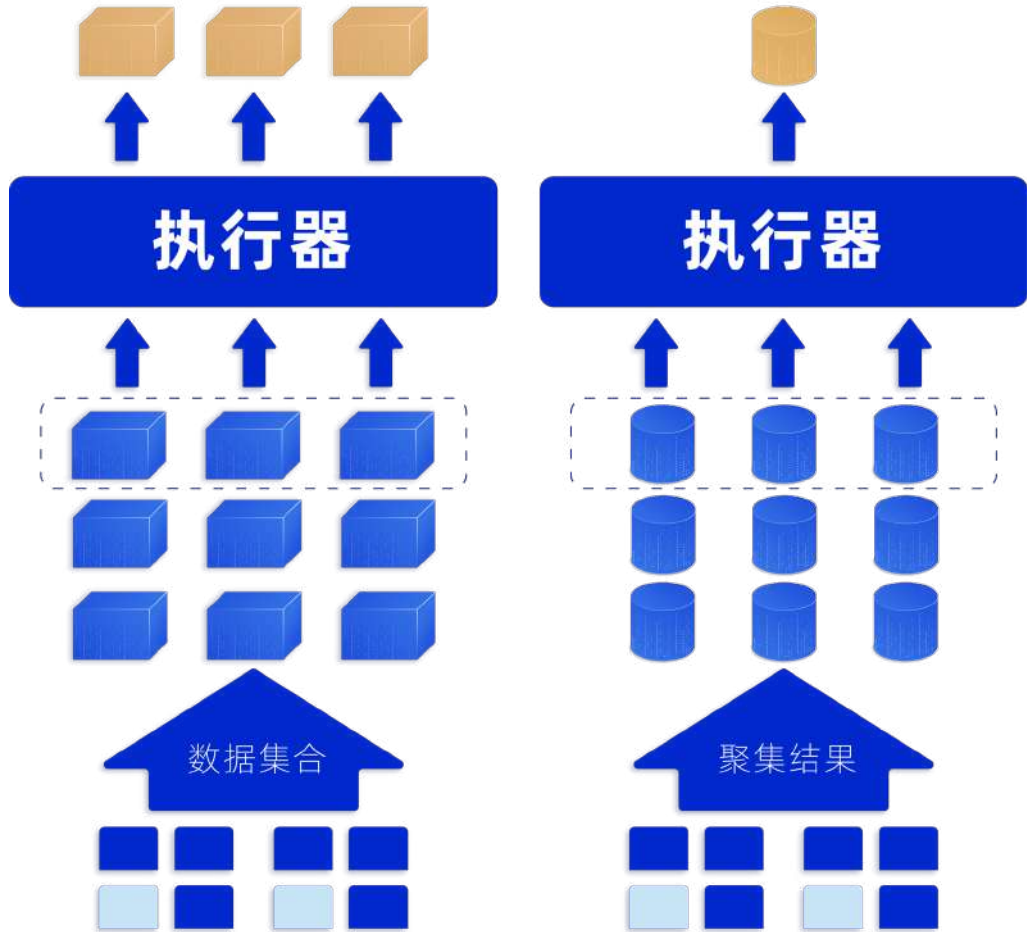
构建新一代云原生存储引擎

- 用户成本（存储成本）
 - 自动选取适应类型的编码
 - 压缩
 - 减少对象存储的访问开销
- OLAP 性能
 - 多级缓存
 - 行列混合存储
 - 定义内外存的数据格式
 - 文件内统计信息
 - 智能Analyze



构建新一代云原生存储引擎

- 完备的事务
 - Block文件级别的MVCC实现
- 优化器与执行器的演进
 - 向量化
 - 文件查询裁剪 (Block Skipping)
 - 聚集下推扫描 (PreAgg Pushdown Scan)



04 | 优化器

PieCloudDB Optimizer

PieCloudDB Optimizer 是一个基于eMPP架构的云原生分布式优化器，它可以为海量数据集上的复杂OLAP查询提供最优的查询计划。

- 分布式优化器
- 处理复杂OLAP查询
- 云原生优化器

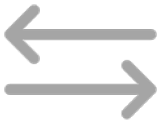
分布式优化器



处理复杂OLAP查询



多表连接的最
优顺序搜索



多阶段聚集



分区表的静态
和动态裁剪



相关子查询的
提升转换



CTE和递归CTE
的优化




等等

云原生优化器


针对云环境的特性，提供更多高阶的优化



聚集下推



预计算



文件剪裁

PART 04

PieCloudDB 2.1新特性

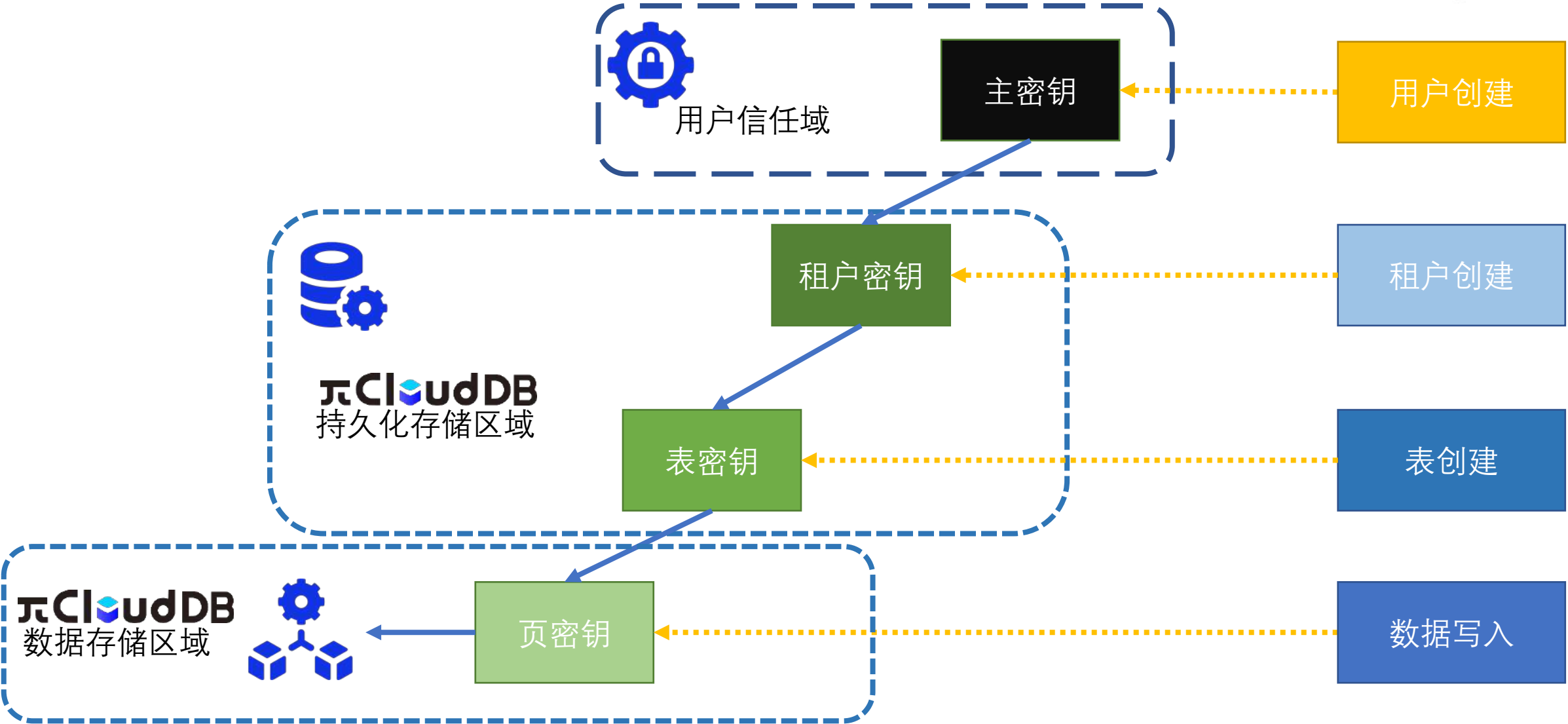
安全性增强

全链路优化

生态建设

安全性增强

- 透明加密技术
 - 加密用户数据，避免被未经许可人员读出
 - 用户无感知，不影响用户的业务，对性能影响小
- 合规
 - 符合数据安全审计要求
 - 符合业务安全审计要求



安全性增强

- 云原生安全
 - 传输层加密
 - 缓存数据加密
- 存储安全
 - 元数据持久化存储
 - 用户数据多副本加密储存
- 计算安全
 - 集群失效不影响用户数据
 - ACID保证

全链路优化

- 全新的存储引擎简墨 (JANM)
 - 基于对象存储的行列混存架构
 - 压缩比更好
 - Cache命中率更高
 - 降低CPU使用率



全链路优化

- 高效的分布式优化器
 - 聚集下推
 - 预计算
 - Block Skipping

生态建设

- 更多的云平台的支持
- FDW
- Apache MADLib
- PostGIS

PART 05

总结

PieCloudDB 核心技术优势

OpenPie

- ✓ 首创eMPP分布式技术实现云上弹性大规模并行计算
- ✓ 以云计算架构为设计基础 实现云上存算分离

存算分离

云上计算资源可弹性分配，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。

多云部署

可根据客户需求在任何IaaS云和裸硬件上安装。可打通多云的数据管道，解锁对特定IaaS云的依赖并获得云资源议价权。

数据安全

PieCloudDB提供企业级透明数据加密。运用实时加密，高强度算法，多级密钥等技术保护数据安全。

弹性计算

企业可灵活进行扩缩容，随着负载的变化实现高效的伸缩，轻松应对PB级海量数据。

实时处理

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。

- eMPP: *elastic Massive Parallel Processing* 弹性大规模并行计算



关注OpenPie公众号

获得更多资讯



加入PieCloudDB技术群

了解更多干货



THANKS

Data computing for New Discoveries

数 据 计 算 ， 只 为 新 发 现