

PieCloudDB Database

基于 eMPP

(弹性大规模并行计算) 的云原生虚拟数仓

产品白皮书

目 录

行业背景	3
数据量的爆发式增长	3
数据库的未来在云上	3
传统数仓的痛点	4
云时代的数据处理要求	5
PieCloudDB, 云原生虚拟数仓	6
PieCloudDB 产品概述	7
PieCloudDB 产品架构	7
PieCloudDB 产品特性	8
PieCloudDB 产品核心技术	11
PieCloudDB 产品优势	13
关于OpenPie	15
附录：术语表	16

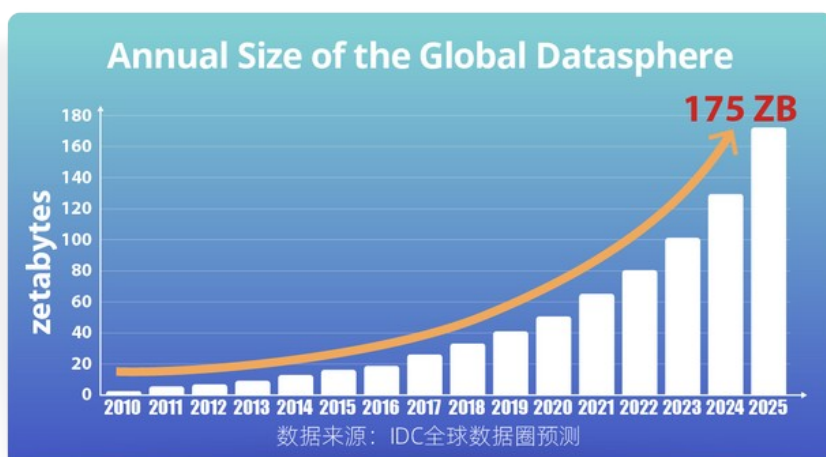


01

行业背景

石油是工业的血液，数据是数字经济的“石油”，数据分析则是石油精炼。

随着信息技术的发展，互联网应用的加速普及，人类进入了数字经济时代。进入二十一世纪以后，随着移动互联网技术、物联网技术、5G等技术的发展，全球数据圈（Global Datasphere）呈指数级递增，IDC预测全球数据圈将于2025年增长值175ZB，而中国的数据圈有望于2025年爆炸式增长为世界第一^①。数据被称为数字经济时代的“石油”，如同石油驱动了工业化时代的进步，大数据将推动智能化与数字化时代的发展。



IDC:全球数据圈预测

数据量的爆发式增长

为了挖掘数据的价值，企业面临着海量数据的存储与分析需求，业务也面临着更多热点及突发流量所带来的挑战。面对数据计算（Data Computing）的巨大诉求、数据组织的运行成本的急剧增加、数据格式的丰富多样，企业的数字化转型面临巨大挑战，急需一款数据库产品，帮助其最大化利用数据资产，降本增效，进行更智能高效的数据计算。

数据库的未来在云上

随着云计算时代的到来，不仅使得计算成本极大地降低，也提供了无限丰富的计算资源，释放出数据计算产生智能的更多机会。早在2019年，Gartner便做出预测：数据库市场的未来在云上^②。在2022年首次发布的《数据库中国市场指南》（Market Guide for DBMS, China）中，Gartner指出，中国数据库行业将加速增长并逐步向云端迁移，未来四年，中国数据库行业向公有云迁移的速度将超过全球平均水平。云原生数据库成为大势所趋，各个企业也都在向这一趋势靠拢。2020年数据显示，云数据库已占据整体数据库市场份额的40%，2022年云数据库营收数据将占据数据库整体市场的半数以上。



Gartner: 数据库中国市场指南

传统数仓的痛点

很多受欢迎的数据库仓库均为分布式数据库，而典型的传统分布式数据库系统大多是 MPP（大规模并行计算）架构。MPP 架构的数据库以 PC 服务器为单位，通过如下图所示的组群方式来扩展存储和计算。假设一个宽表有 3 亿条记录，MPP 数据库会尝试在每台 PC 服务器的硬盘上分布 1 亿条记录。数据计算时，所有机器同时并行计算，理论上最高可以把计算时间降低到单机部署的 $1/n$ （ n 为机器数量），节省了海量数据的处理时间。



传统数据仓库架构

然而，随着数据量的不断攀升，企业对数据仓库的要求也越来越高，在使用过程中，传统 MPP 数据库解决方案迎来了一系列的瓶颈：

缺乏弹性

传统数据仓库的计算和存储是紧密耦合的，计算资源和存储资源按某一比例强绑定，因此用户在扩容时，必须同时扩容计算资源和存储资源，在扩缩容、运维、迁移上都存在一定的挑战。当企业遇到负载高峰时刻或需要紧急得到某个报表结果时，传统数据仓库无法及时扩资源，导致大数据系统无法弹性、快速地分析业务数据，错失了充分挖掘数据价值所带来的商业机会。

成本高昂

传统数据仓库价格高昂的软硬件、开发运维人员的高昂薪资需要企业进行巨大的前期投入。传统数据仓库客户的生产环境资源利用率，无论是存储或是计算资源往往都不尽人意。随着存储和工作负载需求的日益增长，面临数据库的扩容和升级时，由于传统数据仓库架构存储和计算的紧密耦合，往往需要企业花费巨大的运维和时间成本，且操作繁琐。

木桶效应

传统 MPP 数据仓库架构存在“木桶效应”，集群整体执行速度取决于最“短板的”节点的性能。因此，一个节点的表现往往会“拖垮”整个集群的性能，导致查询速度变慢。随着时间的推移，业务的增长，企业往往需要在1-2年后对集群增加计算节点，此时，无论新的计算节点性能如何好，集群总体性能都会受制于老的节点。因此真实生产环境中，常常见到客户在需要扩容时，采取重新新建集群的方式。

数据孤岛

随着业务的发展，数据量的增加，和信息化建设的需求，企业会为不同部门建设相应的业务信息化系统。我们在真实客户场景中，常常看到很多企业有成百上千个集群，但这些集群的元数据往往都是一样的。这种情况下，很多元数据会在不同集群间存在不一致的版本信息。此外，如果企业需要做跨集群的访问，往往非常困难，会造成数据孤岛的存在。

运维成本

对于传统 MPP 数仓，企业往往会需要配备运维人力，且对运维、开发人员要求高，需要相关人员掌握复杂的技术栈，技术的更新迭代迅速，相关人员需保持积极的知识更新意识。相关人才市场较小，人才匮乏。高昂的学习成本造成用户使用过程中性能差、故障率高、故障修复时间长等问题。

云时代的数据处理要求

随着数据量和计算能力的爆发式增长，云计算技术的迅猛发展，云原生架构愈受欢迎，云原生时代应运而生。云原生时代，越来越多的企业将应用向云上迁移，而越来越多的数据也流向云上。公有云带来了众多优势：

随时可以申请/释放的计算资源

无限的计算资源

无限的存储池

低价的对象存储

这些优势使云原生数据库得以降低计算成本，提供无限丰富的计算资源，实现分钟级的伸缩性和真正的高可用，释放出数据计算产生更多智能的机会。以下是云原生时代数字企业的典型使用场景：

- **每天有数个小的计算任务，需要数个节点**
- **每周有一个中等计算任务，需要数十个节点**
- **每月有一个大的计算任务，需要数千个节点**

面对这些不断变化的业务需求和计算任务，企业产生了更高的需求：

- **无限空间：**能够提供无限存储空间，
- **灵活伸缩：**随时根据业务需求弹性增加集群和工作节点
- **资源回收：**在集群完成计算任务时，可以进行资源回收，节省成本

充分结合云计算、大规模并行处理技术的云原生虚拟数仓 PieCloudDB 应运而生，PieCloudDB 帮助企业摆脱了 PC 架构的限制，满足云原生数字时代需求，成为更好的选择。



杭州拓数派科技发展有限公司（又称“OpenPie”）认为计算技术目前经历了三代平台：①大型机时代；②PC机时代；和③云计算时代。每一代计算平台的变更，都带来了数据计算技术的突破性创新的可能性。随着计算技术从大型机时代变革为PC机时代，PC机逐渐取代大型机，极大地降低计算门槛，计算资源日渐丰富，数据计算技术突破性创新。

OpenPie 以 “Data Computing for New Discoveries” 「数据计算，只为新发现」为使命，旗下云原生虚拟数仓 PieCloudDB，运用全新 eMPP（elastic Massive Parallel Processing）分布式技术，可将物理数仓整合到云原生数据计算平台，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。

PieCloudDB 为企业构建「坚如磐石」的虚拟数仓，以云资源最优化配置实现无限数据计算可能，基于新一代数仓虚拟化，提供云数仓智能化解决方案，助力企业建立以数据资产为核心的竞争壁垒。

PieCloudDB 产品概述

拓数派旗下旗舰产品PieCloudDB，是以对行业顶级数据库的抽象思考和设计原则复用为技术路线，可将物理数仓整合到云原生数据计算平台，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。在云上，数据计算资源按需扩缩容，提升数仓的敏捷性和弹性，助力企业降低数仓管理复杂度，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。PieCloudDB在eMPP分布式专利技术、服务器无感知（Serverless）及TDE（透明数据加密）等多项核心技术加持下，为企业构建高安全，高可靠，高在线「坚如磐石」的云原生虚拟数仓，助力企业实现数据价值最大化，更好地赋能业务发展并走向绿色，成为新一代AI数据计算基础设施的一个典范。

PieCloudDB 产品架构

PieCloudDB 整体架构分为三个层次，包括基础设施层、数据处理层及数据应用层。详细阐述如下：

• 基础设施层

基础设施层为 PieCloudDB 提供计算资源、存储资源和网络资源，PieCloudDB 支持部署在物理服务器、虚拟机以及容器中，同时也提供 PieCloudDB 公有云 SaaS 服务。

• 数据处理层

PieCloudDB 核心服务层，提供了并行数据处理能力，拥有元数据节点、计算节点、存储节点以及云原生管控平台节点等共四种角色，具体说明如下：

- 1. 元数据节点：**提供元数据服务，如元数据存储共享、分布式锁、多版本管理、多集群并发、高可用以及用户权限等功能；
- 2. 计算节点：**无状态节点（包括 Coordinator 和 Executer），主要负责接收用户请求和数据计算，支持动态弹性伸缩，提供数据查询、执行计划、查询优化、数据加载、连接管理、并行计算以及资源隔离等功能；
- 3. 存储节点：**存算分离架构，支持本地存储和云存储，推荐采用对象存储，提供数据压缩、数据加密、多模存储以及多级缓存等功能；
- 4. 云原生管控平台节点：**PieCloudDB 集群管控节点，提供数据洞察和集群运维等功能，支持可视化的数据分析、性能监控、集群启停、自动化部署以及权限管控等能力；

• 数据应用层：

用户或者应用可直接调用 PieCloudDB 云原生虚拟数仓服务进行数据分析，提供标准的 SQL 接口，且内置各种分析工具，并原生兼容 Postgres 生态，可以很好地处理地理信息数据和文本，未来会扩展其他 API 接口，支持常见的数仓的数据分析和人工智能、数据科学等功能。

PieCloudDB 产品特性

• 全面的 SQL 兼容度

PieCloudDB 高度兼容 SQL: 2016 标准，完全支持 SQL: 1992 标准、大部分的 SQL: 1999 和部分 SQL: 2003 标准（主要支持其中的 OLAP 特性），支持窗口函数等进阶表达式。此外，PieCloudDB 兼容 PostgreSQL 协议，支持标准数据库接口（ODBC、JDBC 等）。

对 SQL 的全面支持和多种过程语言（Procedure Language）的支持使得 PieCloudDB 可以无缝集成业内常见的提取/转换/加载（ETL）和 BI（商业智能）工具。企业只需安排少量的集成工作，就可以使用现有的使用标准 SQL 结构和接口的分析工具让应用在 PieCloudDB 上运行，从而避免了企业受制于供应商，帮助企业在控制业务风险的同时推动创新。

• 多维度弹性扩缩容

PieCloudDB 采取存算分离的 eMPP（弹性大规模并行计算）设计架构。存储侧支持标准对象存储，可以充分利用云计算平台的优势，让对象存储接近无限的容量，计算侧在设计上充分考虑无状态实现，计算节点可以充分利用云环境海量的计算节点池，按需扩容和缩容。

PieCloudDB 能够根据业务需求进行横向、纵向以及集群级别的弹性伸缩。用户可以灵活考虑业务和数据量的变化，动态调整 PieCloudDB 集群中计算节点的数量和虚拟数仓的数量，用最合适的资源量来满足其业务需求。

• 高可用能力

PieCloudDB 实现了元数据、计算、存储等三层架构，计算节点不存储用户数据，是无状态的。当计算节点发生故障时，PieCloudDB 会自动快速发现并调度新的节点替代故障节点，同时也会在后台尝试修复故障节点，从而保证 PieCloudDB 服务的高可用性。

PieCloudDB 的存储引擎，支持对象存储、HDFS、NAS 等分布式存储，数据的完整性、一致性和可靠性等由分布式存储的多副本、EC 以及灾备能力进行保障，将由于人为或自然原因导致的数据丢失概率降低到最低，保证单个硬件设备的故障不会影响业务。用户可以根据实际需求选用最适合的分布式存储方案。

此外，PieCloudDB 提供的独立的元数据服务（Catalog Service）保证元数据存取的效率和高可用性。PieCloudDB 元数据采用分布式 KV 存储管理，具有完备的高可用方案。在 PieCloudDB 中，每份元数据都将以多副本的形式分散到多个服务节点，并支持定时备份，以确保避免因用户数据的丢失而造成的损失。

• 高性能的数仓引擎

PieCloudDB 采用高效并行的方式进行数据加载和处理，处理速度随节点增加而提升，支持流数据快速加载。PieCloudDB 的 eMPP（弹性大规模并行计算）架构让数据计算可以自动化弹性伸缩，用户可以根据计算任务灵活的分配最优数量的节点执行查询。同时 PieCloudDB 的多集群能力可轻松面对高并发场景，企业可以根据业务的并发需求动态扩展 PieCloudDB 集群，满足业务应用的并发需求。

PieCloudDB 实现了全自动 Analyze，可以在数据发生变化的时候自动完成Analyze，及时生成准确的统计信息。

PieCloudDB 支持元数据和用户数据的缓存，计算节点在执行查询时不再需要频繁地访问元数据服务和存储服务，降低了元数据访问的网络延迟，加快了用户数据的访问速度。

PieCloudDB 对数据查询的执行流程在查询执行器中进行了全链路的优化，完整支持 TPC-H 和 TPC-DS 等测试，实现了包括聚集下推、Block Skipping 等功能模块，大大缩短执行时间，高效的数据查询提高了数据分析的实时性。

1. 聚集下推

在执行聚集函数的过程中，查询优化器会把聚集操作下推到连接操作之前去执行，可以极大地减少连接操作需要处理的数据量，使得查询性能显著提升。在很多分析型场景下，聚集下推会取得百倍或千倍的性能提升。

2. Block Skipping

PieCloudDB 实现了Block Skipping的优化机制，在数据库运行查询语句时，通过预计算每个块(block)中列聚集信息，在执行期间跳过非必要的数据块，减少数据读取量提高查询性能。此外，PieCloudDB PieCloudDB 还支持查询优化器Orca。Orca是一款开源的、基于Cascades 模型的模块化查询优化器，可以帮助用户对SQL进行优化，生成高效的查询计划。

此外，PieCloudDB 兼容 ORCA 优化器（ORCA 是一款开源的、基于 Cascades 模型的模块化查询优化器，可以帮助用户对 SQL 进行优化，生成高效的查询计划）。

• 原生多租户支持

PieCloudDB 通过独特的设计，具备原生多租户能力，每个租户有自己的数据库和虚拟数仓。不同的租户互相隔离，租户之间数据和元数据彼此都不可见。每个租户可以按需创建不同得角色和用户，不同用户可以拥有不同的权限。

• 人性化的管控平台

PieCloudDB 支持多计算集群（也叫虚拟数仓）配置，提供了一键部署功能，可以按需定义规格（CPU/MEM/节点数量等），管理运行状态，在允许租户复用资源的同时保证资源隔离。

PieCloudDB 具备的可视化管理功能界面，支持多种统计数据的汇总和明细展示。提供监报告警功能，具备元数据、计算节点以及平台工具等完善的监控指标体系，同时支持日志收集和展示，方便用户更好地进行智能化运维。

PieCloudDB 管控平台提供数据洞察、元数据浏览、用户管理、权限管理、SQL 查询历史、ETL 管理等功能。

• 湖仓一体化分析

PieCloudDB 支持统一的资源管控，通过云原生存储引擎支持多种存储模型，用分布式计算引擎和数据接口来实现湖仓一体化架构。通过云原生大数据服务以更低成本满足企业业务高弹性和敏捷性的需求。

PieCloudDB 打破了数据湖与数据仓库割裂的体系，架构上实现了将数据湖的灵活性、生态丰富与数据仓库的企业级能力充分融合，帮助企业构建数据湖和数据仓库融合的数据管理平台。PieCloudDB 内置的 foreign-data wrapper (FDW) 模块，让用户可以访问 HDFS、MySQL、Oracle 等外部数据。PieCloudDB 支持 JSON、XML、Parquet 以及二进制数据等格式，同时，PieCloudDB 既支持用户利用过程语言自行开发模块进行数据分析也原生兼容开源机器学习库 Apache MADlib，从而可以原生实现一些高级机器学习功能。该拓展提供数学、统计学以及机器学习方法，包括但不限于线性回归、关联规则、贝叶斯分类、决策树和随机森林等算法支持。同时 OpenPie 团队经验丰富的数据科学家团队可以为企业用户提供相关建议。

• 多云部署

PieCloudDB 可根据客户需求在任何 IaaS（裸金属服务器、虚拟机、K8S 容器等）上运行。借助这种“不受限于基础架构”的跨云部署，企业可以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。企业在 PieCloudDB 上部署时，可以：

- 1. 消除平台/供应商制约：**湖企业可根据自身需求选择不同的供应商来获得 PieCloudDB 的服务与支持；
- 2. 在任何位置运行：**PieCloudDB 不受限于基础架构，企业可自由根据自身需求在任何位置运行，
- 3. 灵活部署：**PieCloudDB 提供了一键部署与扩容选项，企业可根据所需数据节点数量，按需配置。



- **数据隐私安全和加密**

PieCloudDB 提供企业级数据透明加密。实时加密（on-the-fly）、高强度算法、多级密钥、传输加密等技术为企业数据的安全性保驾护航。

PieCloudDB 还支持包括数据库、表级别授权管理等完善的安全及权限管理，帮助企业系统地管理表级别的权限。支持函数、参数等相关的白名单配置，避免用户对系统进行误操作。

此外，PieCloudDB 提供的可视化平台，帮助企业更方便地梳理维护角色授权。安全性上支持端到端的访问控制管理和证书管理。

- **完善的数据生态**

PieCloudDB 兼容 PostgreSQL/Greenplum 生态组件，兼容大部分现有的 PostgreSQL/Greenplum 客户端，商务智能（BI）和 ETL 组件。并且，PieCloudDB 正在加速构建更加完善的数据生态体系，通过创建生态工具、建立合作伙伴生态网络、打造活跃的技术和用户社区等举措，为用户带来更便捷的使用体验。

PieCloudDB 产品核心技术

PieCloudDB 突破了PC时代计算平台的限制，大胆探索云上计算平台的新可能。其云原生架构运用全新 eMPP 分布式技术，在传统 MPP 数据仓库功能丰富和成熟稳定等优势的基础上，加入了众多的云原生特性。

- **多集群数据共享**

PieCloudDB 通过统一元数据管理技术安全轻松地实现不同集群间的数据共享，使数据库中任何一个集群可以对数据库内任何一份数据进行访问，避免了数据复制与移动的复杂性。此外，PieCloudDB 的云原生架构支持用户基于同一份数据建立多个集群，每个集群拥有独立的计算资源和独立的权限管控，满足业务的隔离性和安全性需求。

- **多级缓存技术**

在 PieCloudDB 的计算层，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。其中，元数据作为数据库系统中最重要的特征信息，实现高速的元数据访问是提高数据库查询性能的有效途径。为了避免网络延迟，加速查询优化，PieCloudDB 实现了元数据层全新的缓存架构，有效减少了访问元数据服务器带来的网络通信开销和元数据服务

器的负载，大幅提升了元数据访问的速度，从而提高数据库系统性能。

PieCloudDB 打造了用户数据统一缓存管理功能，使得数据计算在多数情况下均为本地计算，避免了 PieCloudDB 计算节点与对象存储之间的数据移动，提高计算效率，减少网络延迟和系统响应时间，保证用户的实时性需求。

• 全新的存储引擎「简墨」

PieCloudDB 打造了全新的存储引擎--简墨（JANM），实现了基于对象存储的行列混存结构。行列混存结合了行存和列存的优势，允许面向列的压缩方案，压缩率更高，节省存储空间；跳过不必要的列的扫描，提高查询效率；在读取少量数据时，提高Cache命中率，减少 I/O 次数。

PieCloudDB实现了存储中立，支持公有云、私有云、混合云。PieCloudDB 除支持自己的存储格式，还支持部署在存储系统的开源格式，例如 Parquet。

此外，PieCloudDB 针对对象存储进行了优化，对分布式处理和写处理进行了增强，在特定查询场景下，避免了重组操作，提高了 I/O 效率。

• eMPP 分布式技术

OpenPie 发明了云原生虚拟数仓的 eMPP（elastic Massive Parallel Processing）分布式专利技术，在云上，PieCloudDB 利用 eMPP 架构，实现多集群并发执行任务。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。

• 全新的优化器「达奇」

基于自主研发的高效优化器--达奇，PieCloudDB 可以更智能高效地生成统计信息，并生成更高效的查询计划，达奇优化器支持聚集下推，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。

PieCloudDB 产品优势

作为一款云原生虚拟数仓，PieCloudDB 具备的弹性灵活、高性能、稳定成熟等特点，兼备云原生数据库所特有的即开即用、敏捷灵活、安全可靠等优势，为企业的 PB 级海量数据分析需求提供极高的可用性和可靠性。



PieCloudDB 云原生特性

• 弹性灵活

传统数据仓库的计算和存储是紧密耦合的，计算资源和存储资源按某一比例强绑定，为了确保查询结果的准确性，每个计算节点都需参与每条查询的执行中，在扩容、运维、迁移上都存在一定的挑战。企业业务发展的不确定性，导致大数据系统无法及时分析业务数据，错失了充分挖掘数据价值所带来的商业机会。

PieCloudDB 云原生虚拟数仓实现了计算和存储分离，避免了资源的浪费。企业可根据对资源的需求，灵活地以低成本和高效的方式，单独地进行存储或计算资源的弹性扩展，提高了资源的利用率，节省空间成本和能耗开销。

数据计算资源按需扩缩容，实现计算资源配置最优化，提升数仓的敏捷性和弹性，打开无限数据计算空间，更好地赋能业务发展并走向绿色。

- **即开即用**

传统数据仓库往往需要巨大的人力投入进行数据库的安装和调试，PieCloudDB 云原生虚拟数仓的“即开即用”特性为企业节省了大量运维开支。由于 PieCloudDB 计算节点部署于云端，摆脱了物理限制和潜在的延迟，可随时随地通过互联网轻松管理，无需任何硬件。数据随时随地可用，无需处理任何后端技术问题，为企业进行跨部门、跨区域的数据共享和协作开辟了捷径，保证了企业的全球化进程。

- **坚如磐石（高安全、高可靠）**

传统数据仓库将文件和资源存储在同一主机中，以主备节点数据方式补偿节点宕机时间，严重影响数据时效性，增加了运维的成本和难度。PieCloudDB 云原生虚拟数仓具有故障自动单点切换，自动化容灾与高可用能力，避免了单点失败对业务的影响，保证了数据的安全。当业务有升级或更换服务需求时，可对节点进行不中断服务的逐渐升级。

数据透明加密（TDE）技术保证了所有数据在落盘前完成加密，服务器无感知技术 (Serverless) 利用云上无限计算资源和弹性保证了虚拟数仓永远在线可用，S3 存储和跨云灾备能力保证了永不丢数。

- **敏捷可靠**

传统数据仓库的黑箱状态，导致企业无法及时处理扩容、节点故障等问题。而 PieCloudDB 云原生虚拟数仓部署在云上，具备云计算的敏捷、高效特性，且兼具 eMPP（elastic MPP）所特有的弹性、高性能，可以轻松应对 PB 级海量数据，服务之间互相独立，从而对应用或服务提供了更多层的安全保护，也实现了更多的容错服务。其高度自动化的运维工具支持业务应用更加频繁地更新，赋予企业更强大的敏捷迭代能力。

- **降低数仓硬件和管理成本**

将多个数仓归并至云虚拟数仓，打破传统数仓场景下数据孤岛，解决数据多副本问题，帮助企业降低数仓管理复杂度，以更低的成本实现存算资源在云上更灵活的配置。

03 关于OpenPie

杭州拓数派科技发展有限公司（又称“OpenPie”）立足于国内，基础数据计算领域的高科技创新企业。作为国内云上数据和数据计算领域屈指可数的Day-1准独角兽，拓数派致力于在数字原生时代，运用突破性计算理论、独创的云原生虚拟数仓旗舰产品以及之上的算法和数学模型，建立下一代云原生数据平台的前沿标准，驱动企业实现从“软件公司”到“数据公司”再到“数学公司”的持续进阶，加速数字化转型升级。

拓数派旗下旗舰产品PieCloudDB，是以对行业顶级数据库的抽象思考和设计原则复用为技术路线，可将物理数仓整合到云原生数据计算平台，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。在云上，数据计算资源按需扩缩容，提升数仓的敏捷性和弹性，助力企业降低数仓管理复杂度，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。PieCloudDB在eMPP分布式专利技术、服务器无感知（Serverless）及TDE等多项核心技术加持下，为企业构建高安全，高可靠，高在线「坚如磐石」的云原生虚拟数仓，助力企业实现数据价值最大化，更好地赋能业务发展并走向绿色，成为新一代AI数据计算基础设施的一个典范。

术语表

- **MPP:** Massively Parallel Processing, 大规模并行处理;
- **eMPP:** elastic Massively Parallel Processing, 弹性大规模并行处理;
- **虚拟数仓:** Virtual Data Warehouse, 在数据仓库基础之上抽象出来的一种资源调度方法, 它可以整合多个数仓的资源;
- **云原生:** Cloud Native, 云原生技术使组织能够在新式动态环境（如公有云、私有云和混合云）中构建和运行可缩放的应用程序, 能够充分适配和发挥云计算的优势能力;
- **TDE:** Transparent Data Encryption, 透明数据加密, 使用加密密钥或DEK实时进行输入/输出数据加密和解密, 整个过程对应用完全透明, 应用无需感知;
- **服务无感知:** 数据仓库的“Serverless”化, 让用户使用数仓服务时不用关心服务器的架构和存在, 所有底层资源的调度全部由云虚拟数仓后台自动化完成;
- **Analyze :** 收集数据仓库中表内容的统计信息, 并将结果存储在系统表中, 查询优化器使用这些统计信息来确定查询的最有效执行计划;
- **ETL :** Extract-Transform-Load, 用来描述将数据从来源端经过抽取（extract）、转换（transform）、加载（load）至目的端的过程;

参考文献:

- ① The Digitization of the World From Edge to Core, IDC, 2018
- ② <https://www.gartner.com/en/newsroom/press-releases/2019-07-01-gartner-says-the-future-of-the-database-market-is-the>

www.OpenPie.com



关注OpenPie公众号
获得更多资讯



加入PieCloudDB社群
了解更多干货

OpenPie | π CloudDB