

SelectDB 基于 Apache Doris 的现代化实时数据仓库

马如悦

飞轮科技 CEO & Apache Doris PMC

SelectDB: 基于 Apache Doris 的现代化实时数据仓库



实时数据

数据的实时导入与实时存储

极速性能

多种查询负载上都拥有极速性能

弹性架构

弹性架构实现资源灵活扩展

开放集成

开放设计支撑高效系统集成

SelectDB 提供两种产品形态

SelectDB Cloud

全托管的 Apache Doris 公有云服务。
已经上线阿里云、AWS、华为云和腾讯云。

SelectDB Enterprise

自管理的 Apache Doris 企业分发版。
部署在物理机/虚拟机、K8s 或者公有云/
私有云上。

SelectDB 发展历程

企业版

SelectDB Enterprise V1

- 企业级内核：内核基于 Doris 1.1
- 命令行管控：集群创建、删除、升级、扩缩容

SelectDB Enterprise V2

- 企业级内核：内核基于 Doris 1.2
- 内置可视化开发 WebUI
- 可视化管控：集群管理、监控报警、巡检等

SelectDB Enterprise V3

- 企业级内核：内核更新至 Doris 2.x
- 企业级特性：跨集群复制（CCR）
- 可视化管控：Profile优化、Agent改造、日志查看、导入与集成、K8s部署支持
- 数据迁移工具

云服务

SelectDB Cloud V1

- SAAS 版本上线邀测
- 存算分离架构
- 阿里云、AWS、腾讯云、华为云

SelectDB Cloud V2

- 增加可视化开发工具 WebUI
- 配置无套餐化，改为仅需要指定vCPU数即可
- 支持手动弹性伸缩、定时弹性伸缩和自动启停。
- 无感升级、缓存预热

SelectDB Cloud V3

- 全面开放，达到GA
- 内核更新至基于 Apache Doris 2.0
- BYOC产品形态
- Saving Plan
- 上线GCP、亚马逊云科技

2022年H2

2023年H1

2023年H2

SelectDB 云服务

SelectDB Cloud

SelectDB Cloud 存算分离架构带来极致性价比



存算分离与弹性计算

- 共享对象存储与本地高速缓存
- 多计算集群，实现负载隔离
- 集群计算节点弹性扩缩容（手动、分时、自动启停）

极致性价比

- 存储冷热分离，冷数据使用低成本可靠对象存储，平均带来 5x 左右存储成本节省
- 计算节点通过弹性扩缩容，减少计算资源浪费，平均带来 5x 左右计算成本节省

无套餐 Serverless化



弹性扩缩容：手动、定时、自动启停

WH_Guide
ALBJ3PBD

集群

连接

监控

用量

设置

Cluster01 运行中

基础信息

集群 ID

c-q89ynmwe3g8tqah1hd

集群名称

Cluster01

集群缓存空间

200 GB

计费方式

按量计费 转包年包月

其它信息

创建者

guide@selectDB.com

创建时间

2023-06-25 11:44:06

运行时长

0 天 0 小时 4 分钟

手动弹性伸缩 | 分时弹性伸缩

计费方式

按量计费

当前计算资源

4 vCPU, 32 GB 内存

当前缓存空间

200 GB

目标计算资源 (vCPU)

4 256 512 768 1024

4

默认每个集群最多 1024 vCPU，如需更高配额，请 联系我们 申请

当前费用

¥2.42 / 小时

9.7 折 首 ¥0.07 / 小时

确定

WH_Guide
ALBJ3PBD

集群

连接

监控

用量

设置

Cluster01 运行中

基础信息

集群 ID

c-q89ynmwe3g8tqah1hd

集群名称

Cluster01

集群缓存空间

200 GB

计费方式

按量计费 转包年包月

其它信息

创建者

guide@selectDB.com

创建时间

2023-06-25 11:44:06

运行时长

0 天 0 小时 7 分钟

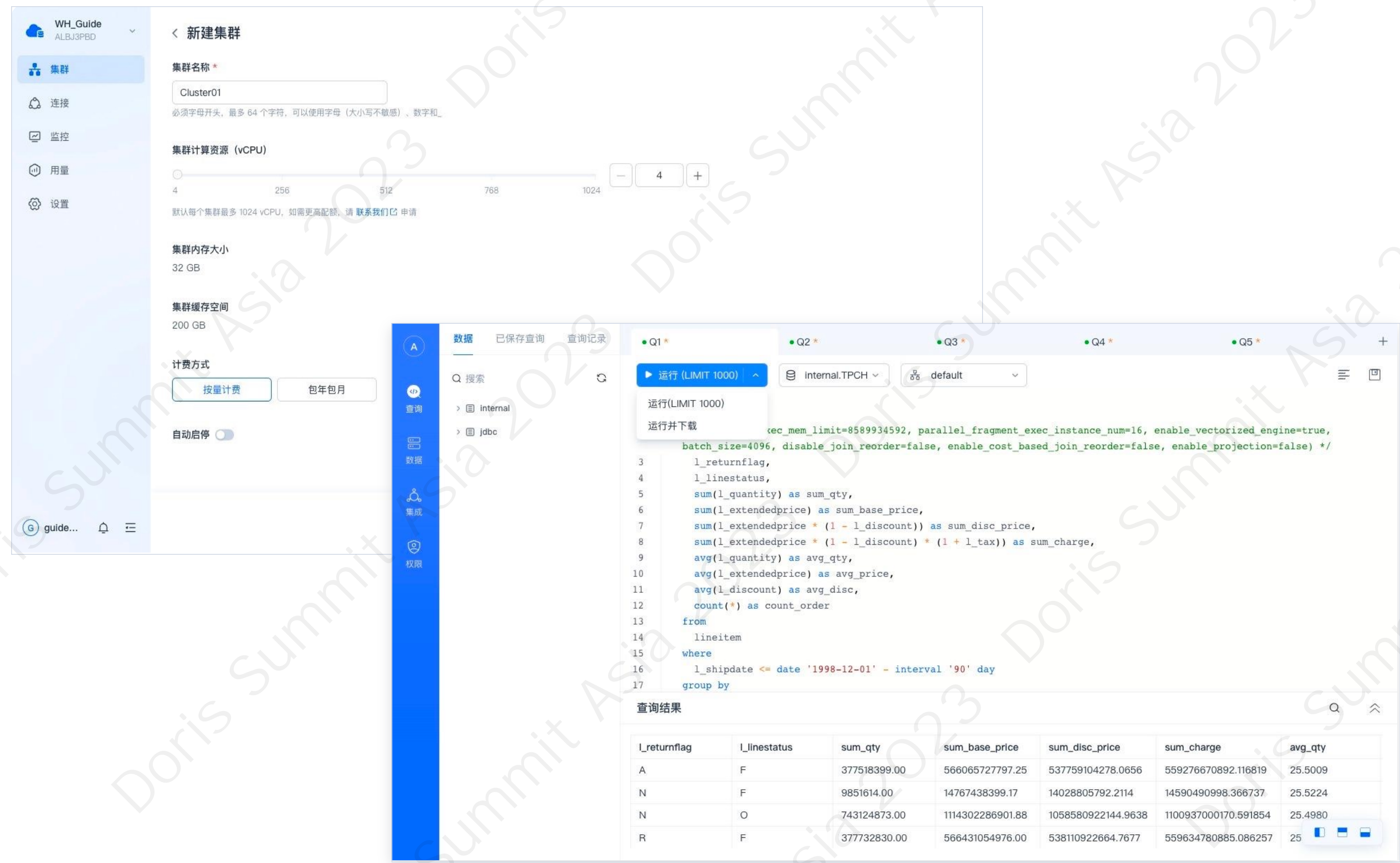
自动启停

自动启停

自动停机触发条件：闲置时长 大于 10 分钟

自动启动触发条件：暂不支持，需要手动启动。

SelectDB Cloud 降低运维复杂度，提升开发效率



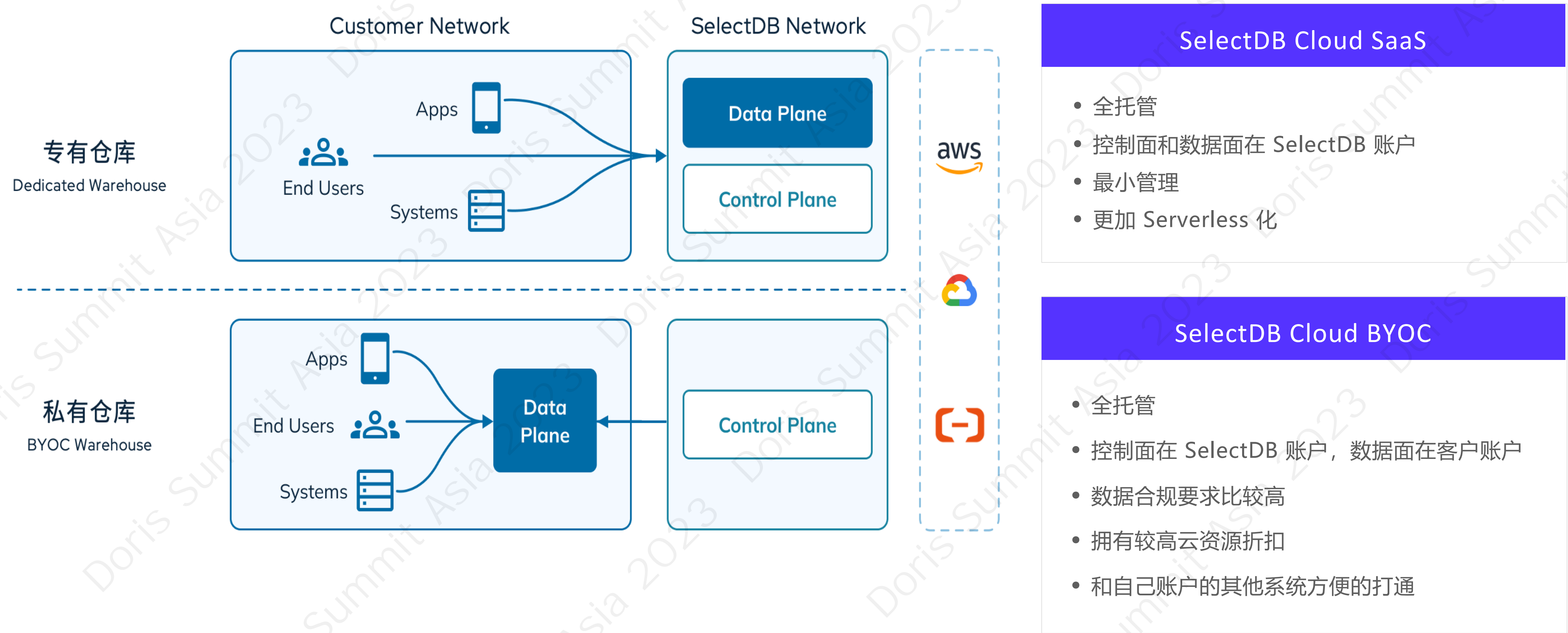
可视化篇理控制台，降低运维复杂度

- 多云统一管理
- 组织管理、计费管理、云市场打通
- 集群管理：Serverless化配置，无需配置节点套餐与数量
- 安全连接：支持公网连接及 IP 白名单设定；支持私网连接
- 监控与报警

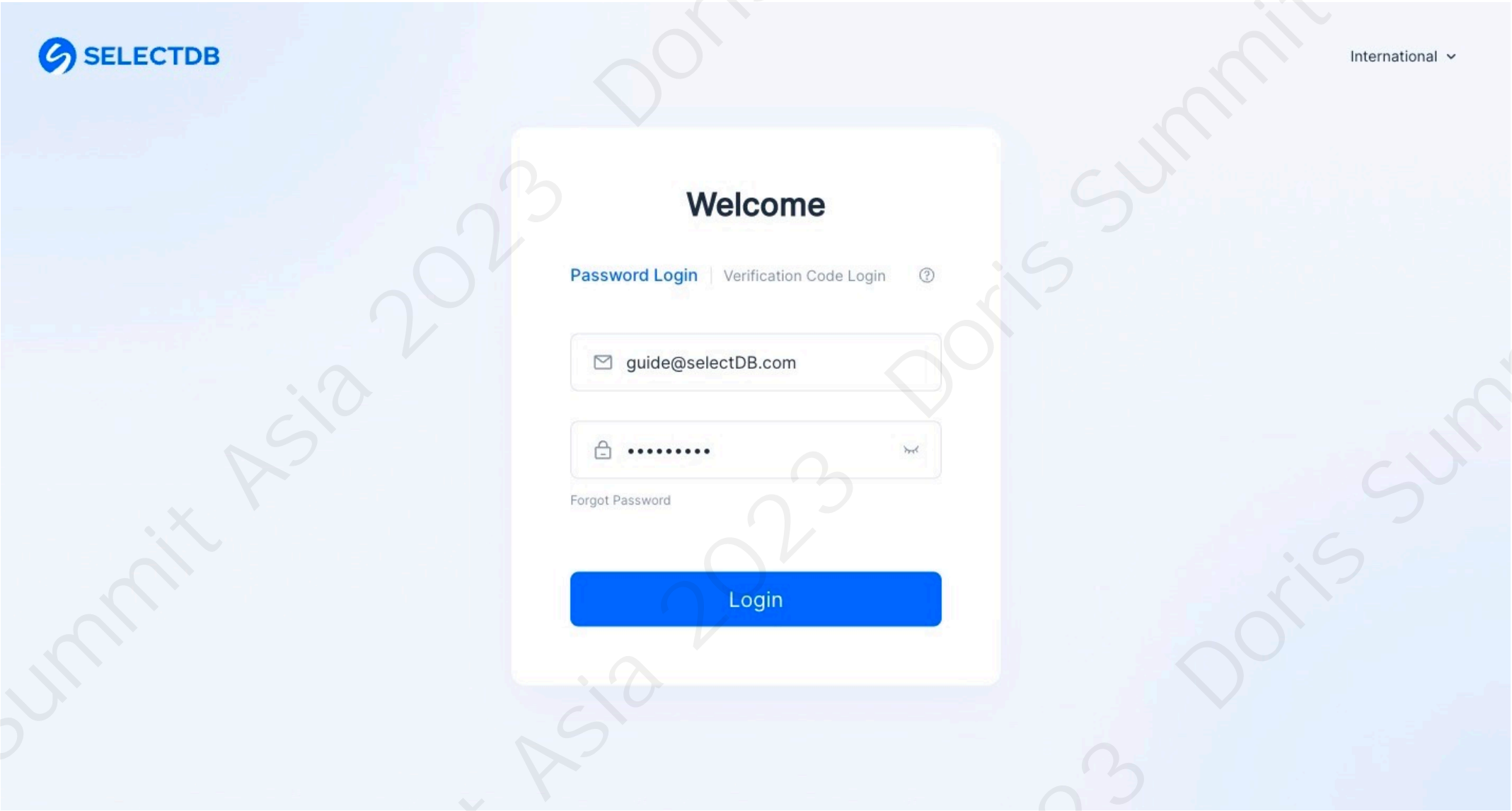
可视化开发 WebUI，提升开发效率

- 数据查询
- 数据集成
- 数据管理
- 权限管理

SelectDB Cloud 两种部署形态



SelectDB Cloud 全面开放，免费试用



国际站

已经上线:



即将上线:



中国站

已经上线:



即将上线:



阿里云 SelectDB 版本邀测版上线

☰

阿里云

资源组 ▾ 华北2 (北京) ▾

🔍 搜索

费用 工单 备案 服务 简体中文

🔔 4

👤

云原生数据库SelectDB

实例列表

实例列表

帮助文档

新建实例

实例ID ▾

输入搜索内容...

🔍

⏴ ⚙️ ⏵

☐	实例ID/名称	状态 ▾	付费方式 ▾	集群个数 1L	创建时间 1L	标签	操作
☐	eirids92-beijing-20180823 实例名称	🔄 创建中	按量付费 转包年包月	0个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🔴 实例锁定	按量付费 转包年包月	2个	2020-08-08 10:00:00	🏷️	详情 删除实例
☑️	eirids92-beijing-20180823 📄 实例名称 📄 🖱️	🟢 运行中	按量付费 转包年包月	4个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🔄 升级中	包月 8 天后到期 转按量付费 续费	9个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🔄 升级中	包月 8 天后到期 转按量付费 续费	2个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🔄 恢复中	按量付费	2个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🔴 实例锁定	按量付费	5个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🟢 运行中	包月 8 天后到期 转按量付费 续费	2个	2020-08-08 10:00:00	🏷️	详情 删除实例
☐	eirids92-beijing-20180823 实例名称	🟢 运行中	按量付费	3个	2020-08-08 10:00:00	🏷️	详情 删除实例

🟡

SelectDB 企业版

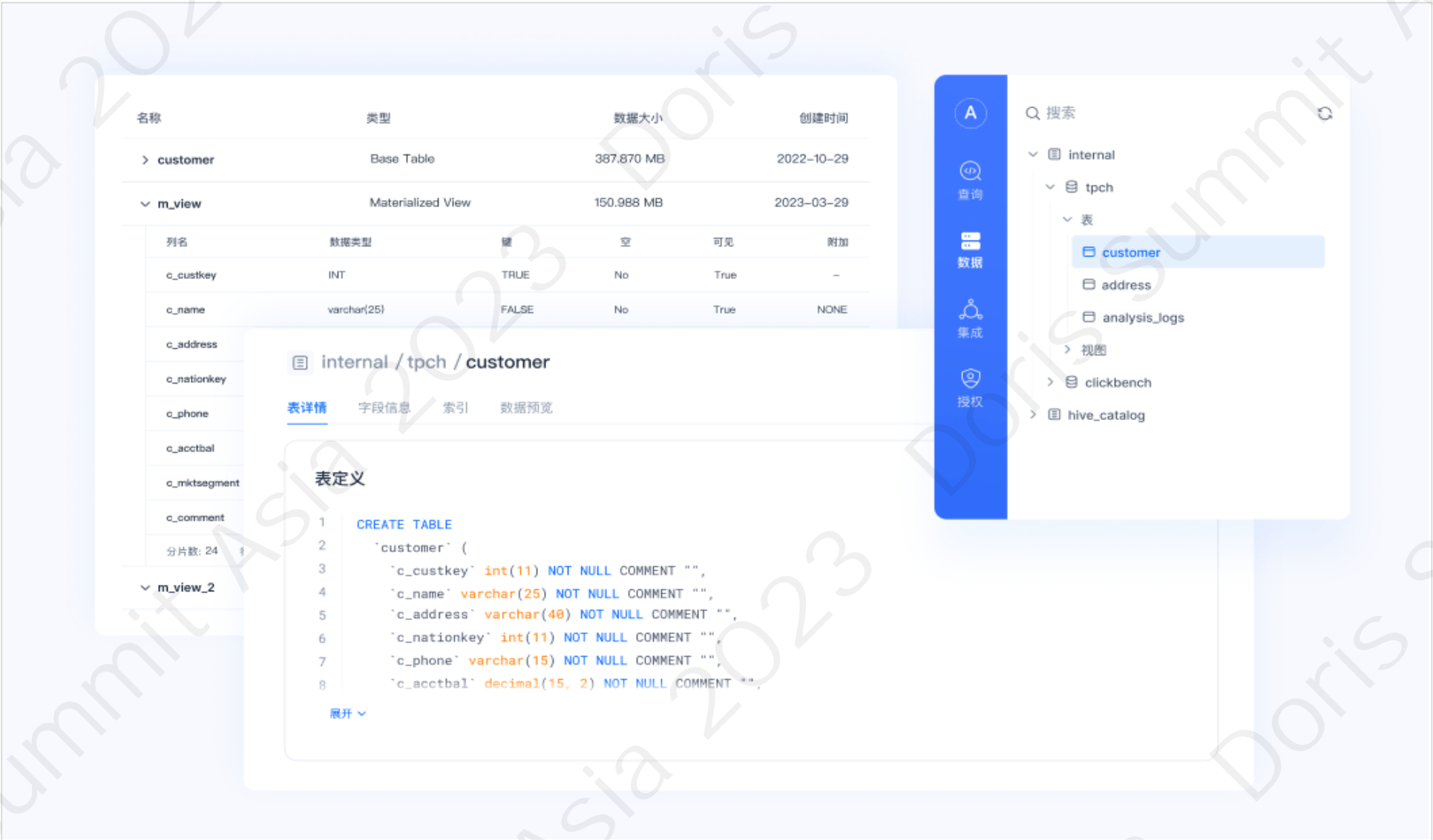
SelectDB Enterprise

Enterprise Core: 与 Apache Doris 100% 兼容的企业版内核



长周期支持的 稳定内核

- 飞轮科技严格测试的稳定内核
- 比社区更快速的缺陷修复
- 提供对大版本的长周期支持



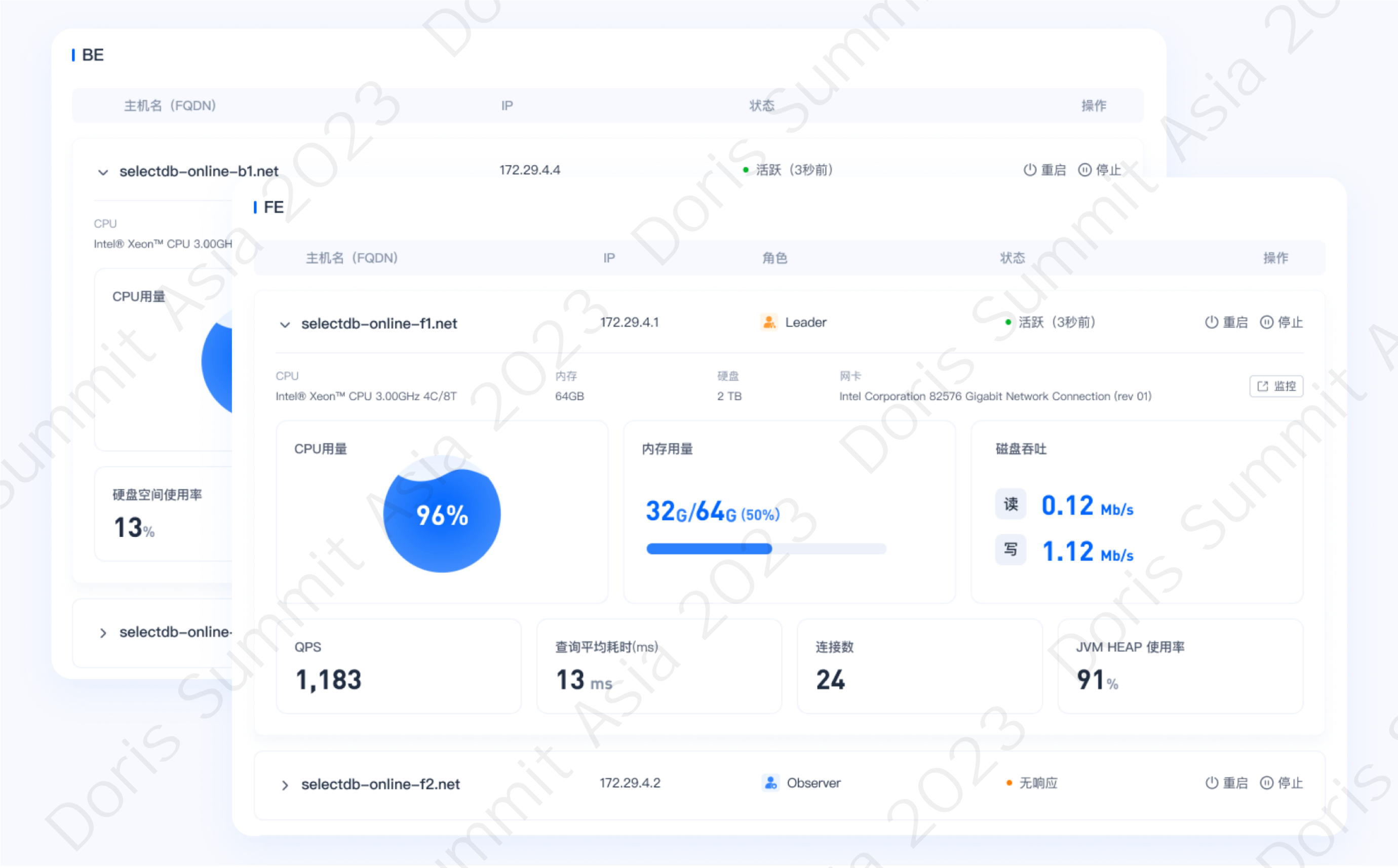
内置可视化开发 WebUI

- 数据查询
- 数据集成
- 数据管理
- 权限管理

Enterprise Manager: 可视化集群篇控工具

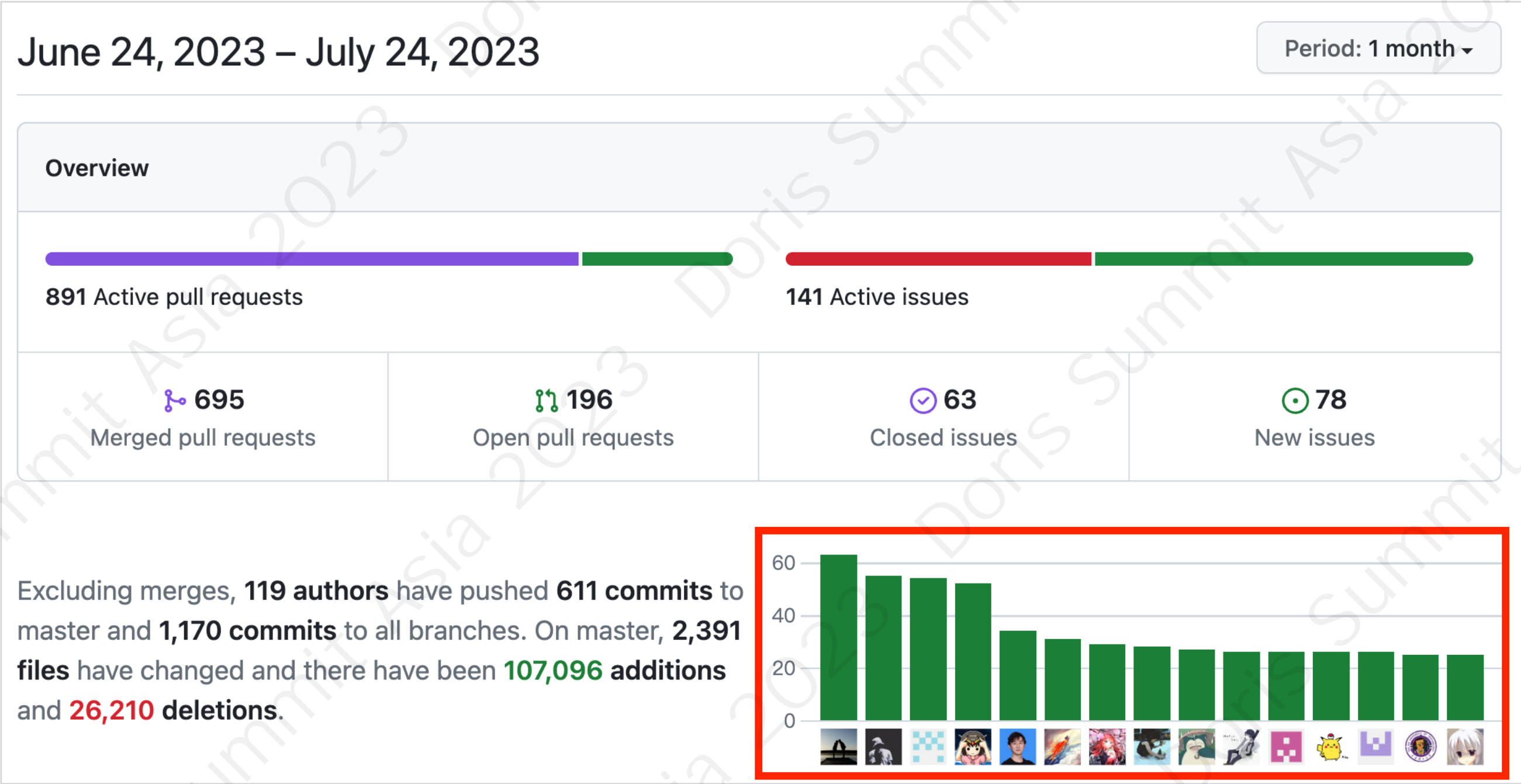
降低运维复杂度，提升运维效率

- 可以管理开源 Doris 内核和企业版内核
- 支持托管已部署集群
- 支持多集群的创建、配置变更、升级、扩缩容等功能
- 提供监控、报警、巡检、审计等功能
- 支持在物理机/虚拟机、 K8S 部署 (开发中) 和公有云部署 (开发中)
- 支持对接客户的私有云



Enterprise Support: 飞轮科技的专家技术服务

月 Top 15 贡献者, 12 名在飞轮科技



- 5 名 PMC 成员, 21 名 Committer
- 主导发布了 Apache Doris V1.0 V1.1 V1.2 V2.0 等一系列版本
- Apache Doris 社区用户群 95% 的问题由飞轮科技同学响应
- 最近两年, Apache Doris 90% 的新增代码由飞轮科技同学贡献

专业团队, 断绝一切后顾之忧

- **消除风险**: 提供例行巡检, 及时消除系统可能的隐患。
- **解决问题**: 严格的服务SLA, 确保 7*24小时及时获得专属支持; 对紧急Bug提供天级别的专属修复版本更新。
- **优化系统**: 通过对产品的使用培训, 业内最佳应用实践的分享, 与客户一起优化系统性能和成本。
- **咨询培训**: 提供对业内最佳实践的咨询、产品的系统培训。

自主可控、安全可靠

- 公安部等保三级，6 项 ISO 安全管理体系认证，2022 信创产业实干者



- 十多项国产化信创兼容认证

麒麟软件

统信UOS

openEuler

Phytium飞腾

Hi6ON 海光

Kunpeng

SelectDB 解决方案

SelectDB Solutions

SelectDB 的主要应用场景

用户行为与画像分析

构建用户数据平台，进行用户参与、留存和转化等行为分析，
及人群洞察和人群圈选等画像分析

日志管理与分析

构建统一的日志管理与分析平台，在极低成本存储的基础上提供
高性能的日志检索与分析

实时报表与实时决策

面向内部和外部的实时报表与仪表盘，面向自动化程序的实时决策

交互式探索分析

对 TB和PB 数据进行快速的即席查询 (Ad-hoc Queries)

混合云和多云部署

无论是混合部署(公有云、私有化部署)还是在多个公共云上部署，统一体验，一切可简化

混合云和多云正在成为用云主流趋势

灵活的产品使用和部署形态

- 开源开放，与 Apache Doris 100% 兼容。
- 提供可以在 IDC、私有云部署的 SelectDB 企业版。
- 提供在全球七个主流云上云原生的 SelectDB 云服务。
- 所有形态都拥有一致的使用体验。

统一的可视化管控工具

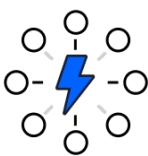
- SelectDB Cloud 提供可视化管控，对多云上的仓库和集群进行统一纳管。
- SelectDB Enterprise 提供可以私有化部署的 Enterprise Manager，可以对部署在物理机/虚拟机、K8s 容器和公有云/私有云上的集群进行统一纳管。

数据贯通和迁移

- 提供跨集群复制能力，可以在多云上进行集群异步复制。
- 提供数据迁移工具，方便从私有化部署迁移到云上，方便在云与云之间迁移。

实时报表与实时决策

实时数据写入



- 支持 100 万行/秒的实时写入，也支持流式的从 OLTP 数据库和 Kafka 同步数据。
- 支持数据更新，关系数据库中某些数据，例如订单状态等，可以实时地反映到数仓中，及时生成有关订单状态等信息的分析报告。

亚秒级查询响应



- 存储和计算引擎全链路向量化，实现数量级的查询加速。
- 强一致聚合物化视图，自动使用预聚合结果，满足亚秒级聚合统计查询。

高并发的解决方案



- 采用分区分桶将单个查询分发到单台机器处理，使用跳数索引 (Zonemap, Bloomfilter) 和点查索引 (主键、倒排) 减少读取数据量，提升并发查询能力。
- 通过点查行存，减少 IOPS 压力，支撑单机上万 QPS 的高并发点查。

稳定、高可用的服务



- 无单点故障，在线扩容、升级自动副本均衡，实现单集群高可用。
- 跨集群实时数据同步，实现跨集群高可用。

从固定报表到交互式即席分析

联邦查询

- SelectDB 支持通过 multi-catalog 的机制将外部的数据库、数据湖映射为 SelectDB 的库表
- 外表和内表可以联邦查询分析。做到了无需移动数据，即可使用 SelectDB 完成多各种数据源的即席查询分析。

多种查询负载上的极速性能

- 极速查询性能：基于成本的优化器，MPP 的分布式执行框架，向量化的计算引擎等
- 对多种查询负载提供极速查询性能：不论是简单的大宽表查询，还是复杂的多表 JOIN 查询，都拥有极速性能。

多计算集群与集群内的负载隔离

- SelectDB Cloud 提供多计算集群能力，Ad-hoc 分析集群和在线服务查询集群共享一份存储数据的同时，做到了集群隔离。
- 计算集群内，可以划分不同的查询资源组，来隔离不同用户的查询。
- 异常 SQL 的捕获：及时发现并 Kill 可能会造成节点内存 OOM 的查询。
- 内存消耗较大的 SQL，会对中间结果进行落盘，减少对内存的消耗。

用户画像与行为分析

基于 SelectDB 建设的 CDP 平台



毫秒级加列、部分列更新

支持行为分析业务属性频繁变动，支持画像场景的宽表列实时更新。

函数支持，业务场景秒级响应

丰富的行为分析函数，带来开发简化和效率提升，比如 retention、window_funnel、sequence_match 等。

高表正交位图实现画像场景的秒级圈人

用户画像千亿数据，秒级人群预估，秒级别 10 标签圈人，10 秒级别 100 标签人群圈选。

日志管理与分析



海量数据实时写入

- 为日志场景优化的高性能倒排索引，写入速度是 ES 倒排索引的 4 倍。
- 服务端 group commit 机制，保证秒级实时可见前提下写入吞吐到 GB/s。

高性价比存储

- 采用列式存储、精简倒排索引、高压缩比压缩，存储空间占用仅需 ES 的 1/5。
- 采用冷热分离，单位存储空间成本仅需 ES 的 1/3，整体成本 1/15。

高效处理业务场景

- 通过分区分桶、时间段筛选，大幅缩小查询的数据范围。
- 通过倒排索引，按检索词快速定位到匹配的日志行，避免大批量扫描，达到秒级响应。

Thank You

谢 谢 观 看



<https://selectdb.com>



获取更多社区动态与最佳实践

Apache Doris 官方平台:

- Apache Doris 官网: doris.apache.org
- Apache Doris GitHub: github.com/apache/doris/

获取更多峰会资料:

- Doris Summit 峰会官网: doris-summit.org.cn
- Doris Summit 峰会回放: <https://space.bilibili.com/1196172099/channel/collectiondetail?sid=1824324>