

阿里云上ClickHouse存算分离架构分享

晓屿 | 产品经理 | 数据库产品与事业部

ClickHouse企业版～阿里云和ClickHouse Inc.独家合作的云原生版本



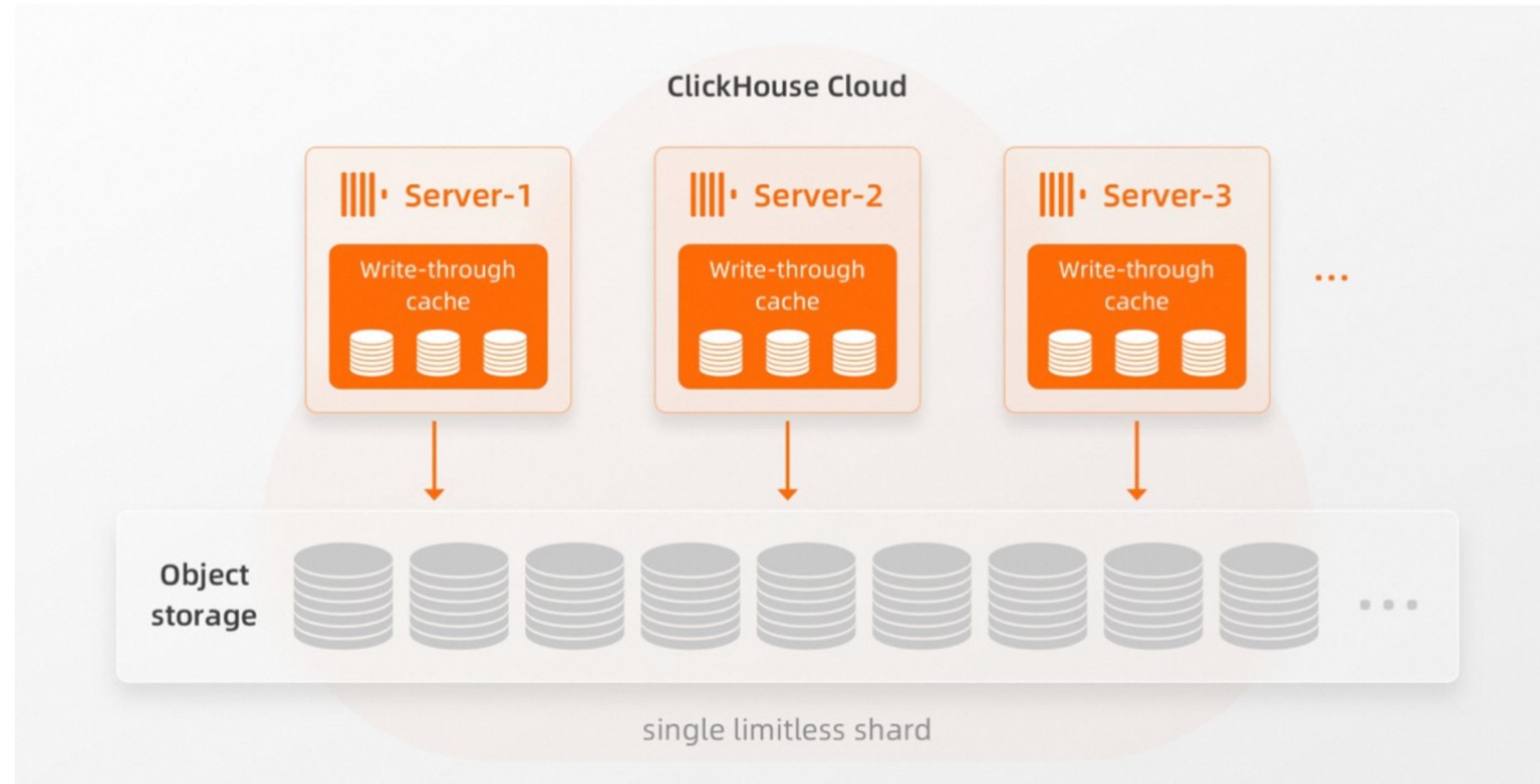
ClickHouse, Inc. and Alibaba Cloud Announce a New Partnership



Nick Peart
Mar 24, 2023

SAN FRANCISCO - March 24, 2023 - Today, ClickHouse, Inc. creators of ClickHouse online analytical processing (OLAP) database, and Alibaba Cloud, the digital technology and intelligence backbone of Alibaba Group, announced a partnership that will enable Alibaba Cloud to offer ClickHouse as an enterprise, first-party service on its platform. This partnership is an exclusive agreement between ClickHouse, Inc. and Alibaba Cloud in mainland China to offer a joint first-party enterprise service in APAC.

clickhouse企业版的存算分离架构



ClickHouse 企业版：缓存层进一步优化查询性能

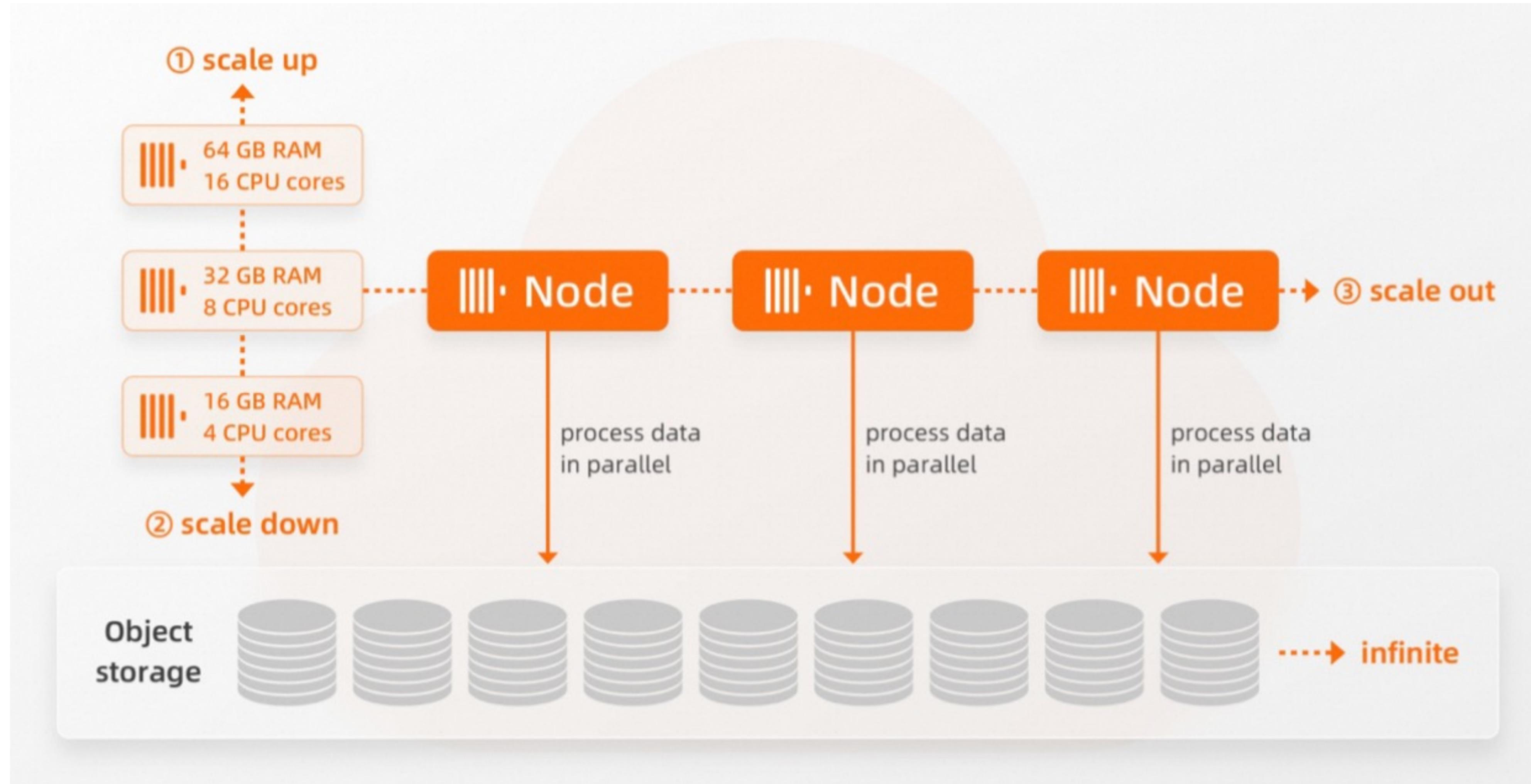
查询性能：

1. 数据集：clickbench标准测试数据集*20倍
2. 测试query集合：clickbench标准query集合

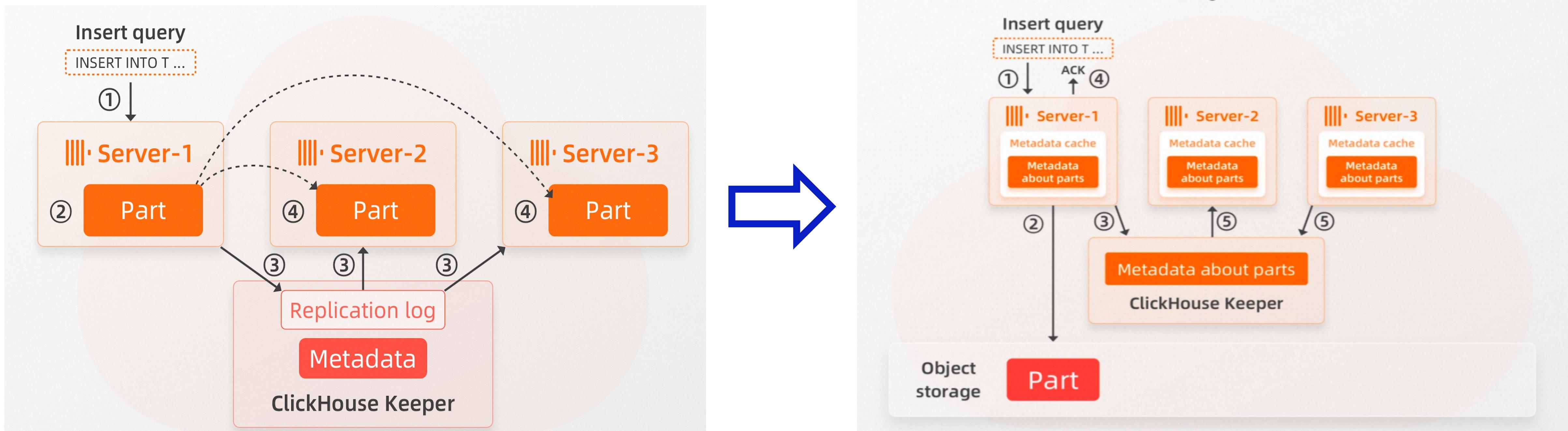
结论：企业版通过增加缓存层，提升了缓存命中情况下的查询性能**1倍左右**



更好的水平扩展效率、更高的容灾能力、更低的容灾成本



ClickHouse企业版：Shared*MergeTree引擎



- Part 数据和计算绑定，数据可靠性弱
- 节点间 Part 数据异步同步，数据一致性弱
- 节点间 Part 数据复制，资源消耗大，效率低
- 存储和计算解耦，计算节点宕机不影响数据可靠性
- 各节点数据同步操作更加轻量和高效，数据一致性强，无资源占用

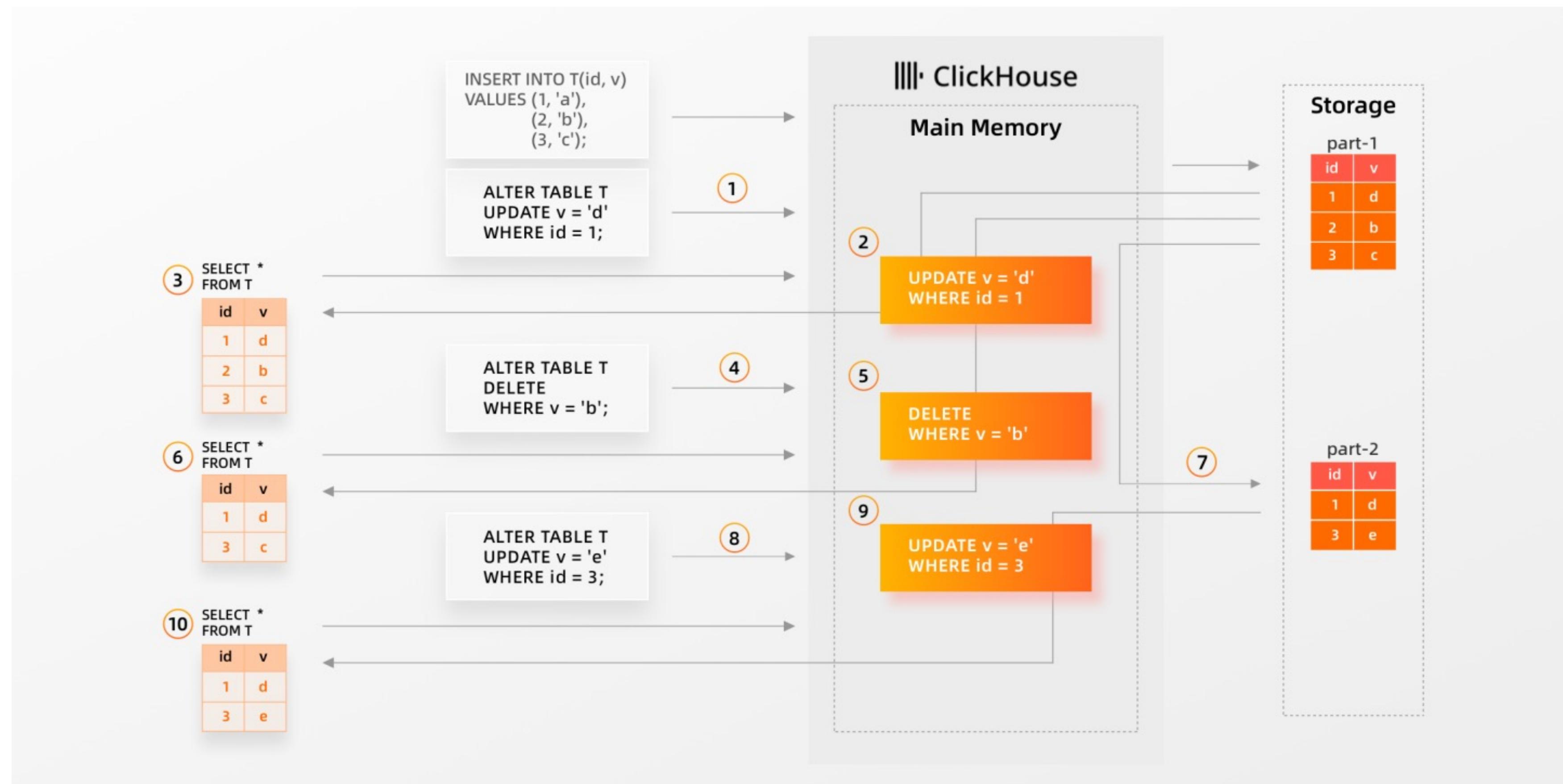
阿里云 ClickHouse 企业版：lightweight update&delete

机制：

1. Server-1 收到写入 update 请求，并添加到队列异步执行
2. 更新操作写入内存，并分发到其他 Server 节点；
3. ClickHouse 收到查询请求，并返回返回更新值；
4. 其他 delete Query 请求发起，添加到队列执行
5. 内存执行 Delete 操作，并分发到其他 Server 节点
6. 返回多个 update & delete 的内存最新查询结果
7. update&delete 批量更新到物理存储
8. 接受处理新的update 请求
9. 和2&5 类似的处理请求

优势：

- 数据结果实时返回，实时性强
- Update & delete 后台多节点并行处理，性能高

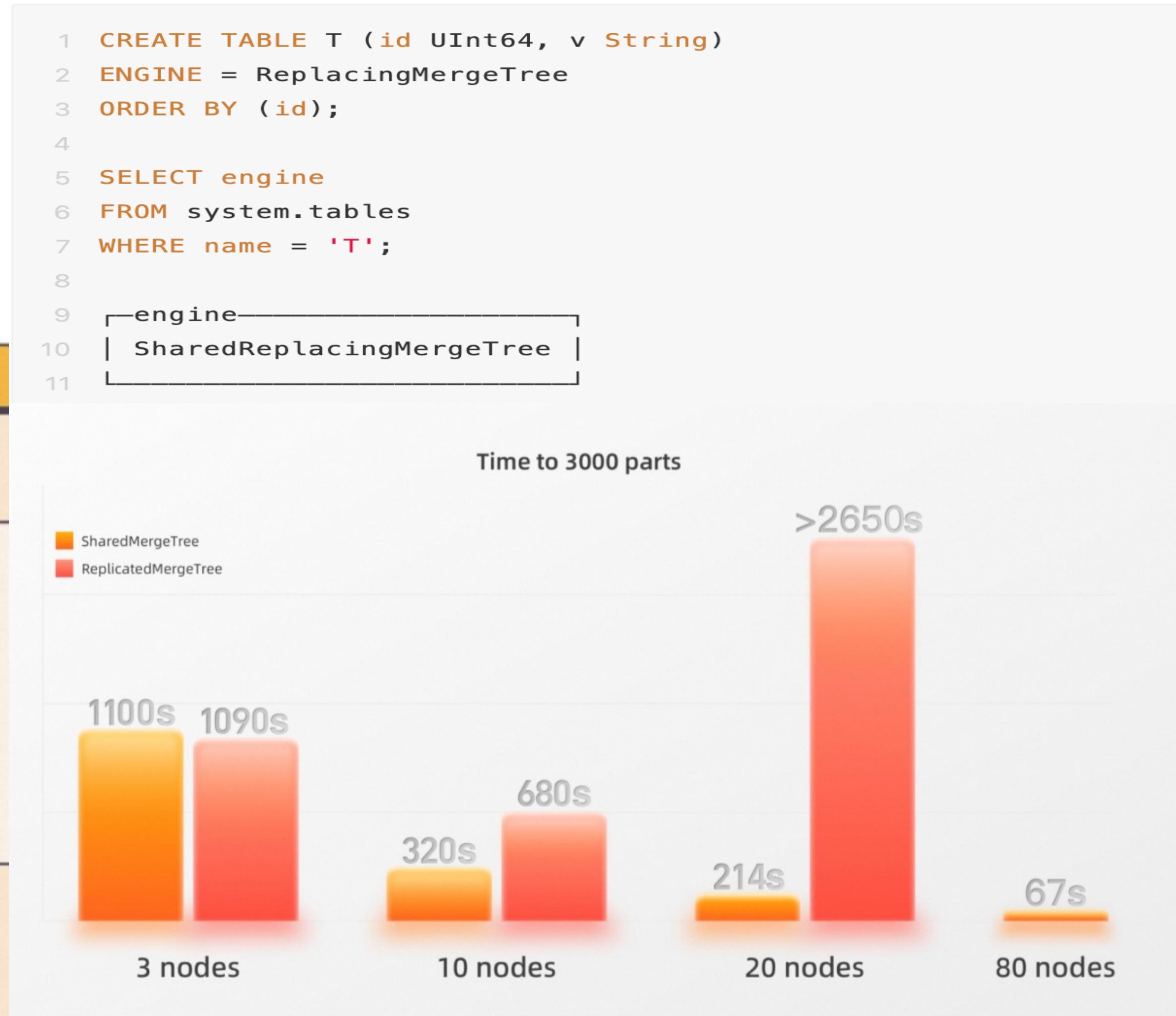


ClickHouse企业版： Shared*MergeTree引擎

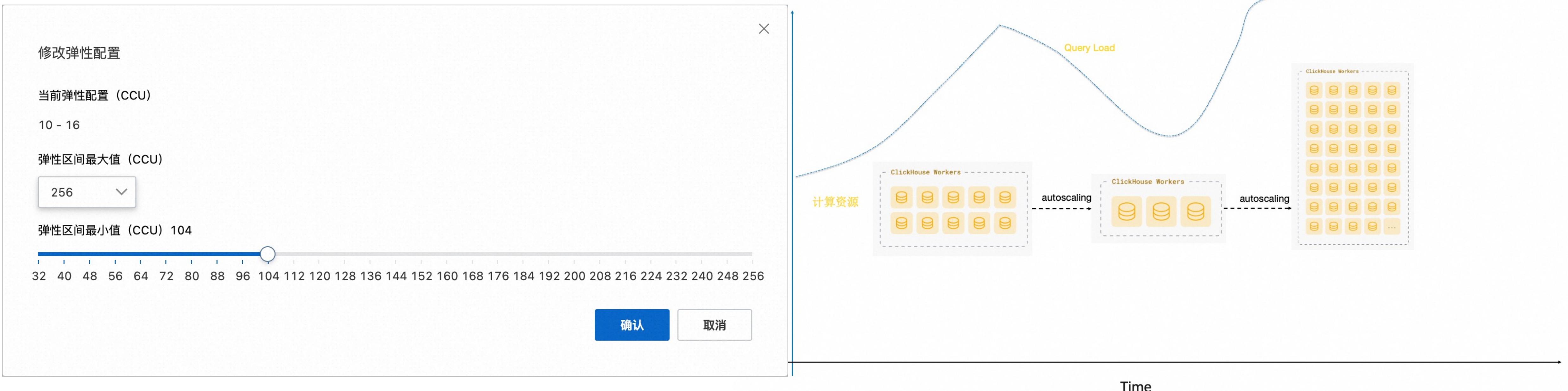
兼容方式：

- *MergeTree 引擎内核自动转换为 Shared*MergeTree 引擎
- 默认高可用，无需使用 Replicated*引擎支持高可用
- 默认分布式扩展，无需使用 Distributed 引擎进行分布式扩展

| | Feature |
|--|---|
| Only on Cloud 仅在阿里云 ClickHouse 企业版支持 | SharedMergeTree Stateless workers |
| Better on Cloud 阿里云 ClickHouse 企业版支持更好 | Lightweight update/delete Replicated Catalog Parallel Replicas Virtual "sharding" keys Optimization for shared storage Multi-level cache Fast backups |
| Both Open Source and on Cloud 企业版和开源版同时支持 | ClickHouse Keeper |

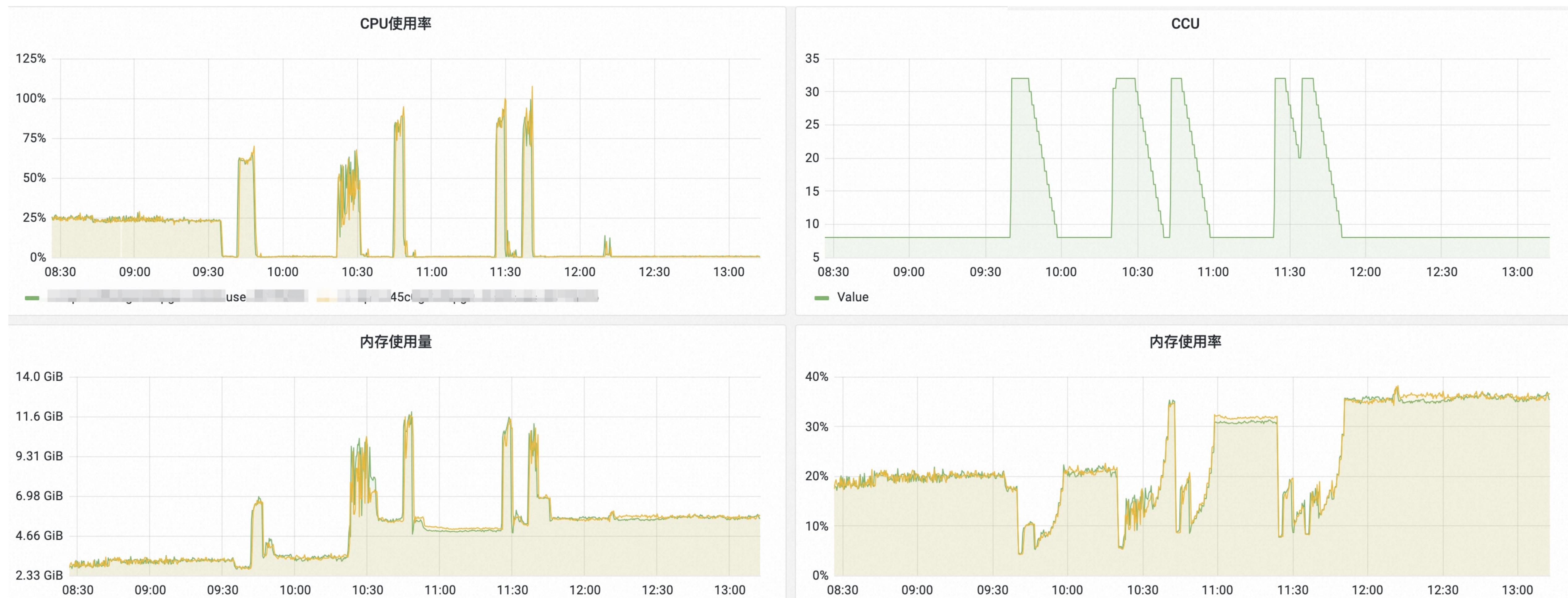


ClickHouse企业版：基于负载的秒级serverless能力



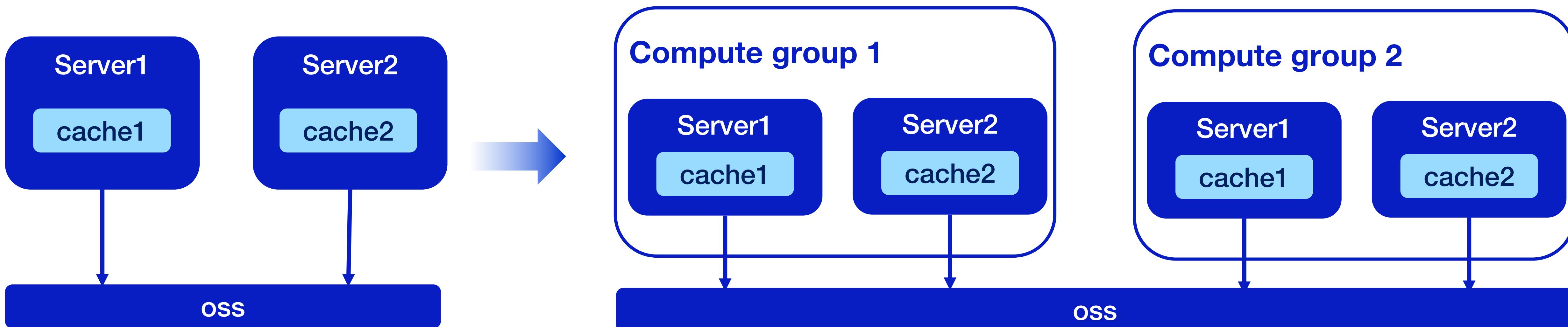
ClickHouse企业版：serverless性能参考

- 1、秒级弹升，跟上业务负载，业务低峰释放资源，Pay-as-You-Go，有效降低企业成本支出。
- 2、无需手动变配，提升业务稳定性，降低运维负担。
- 3、支持保守型和激进型两种弹性策略，用户可根据业务的稳定性和成本诉求自主选择



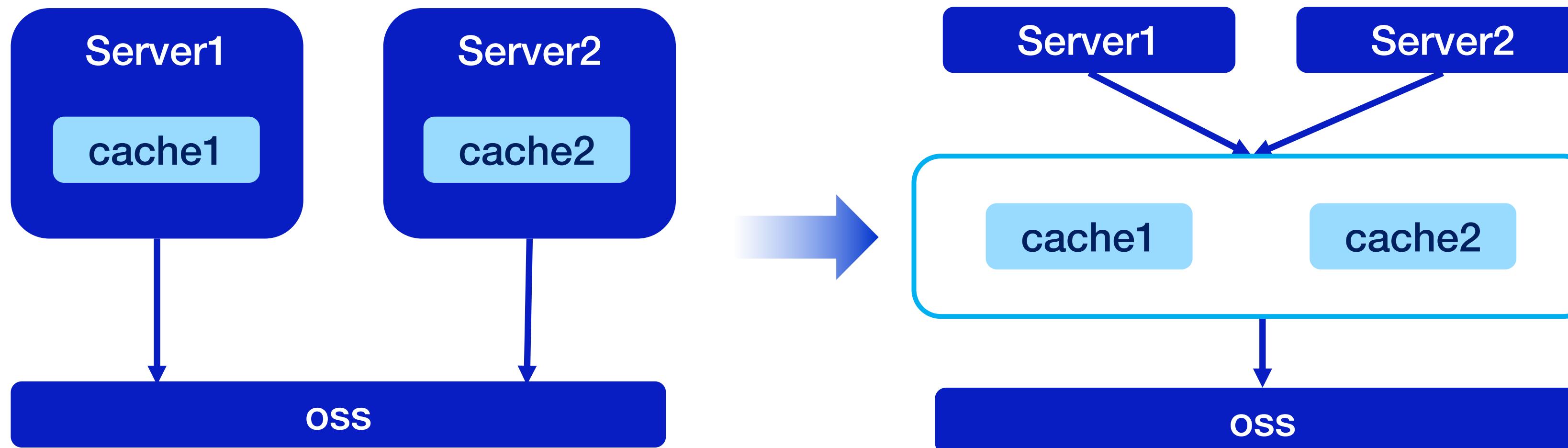
多计算组 (todo)

- 不同业务间，计算资源隔离，避免业务间互相影响
- 不同计算组分别弹性伸缩，灵活控制成本
- 基于多计算组实现的读写分离
- 基于多计算实现的merge和业务流量分离



分布式缓存 (todo)

- 共享存储->共享存储和缓存
- 更好的查询性能
- 更好的新增节点的查询性能保障



按需启停 (todo)

- 存算分离，停机后只针对存储资源进行计费
- 冷热池、镜像预热，实现实例快速创建和启动
- 离线调度：定期的触发资源池内空闲ECS平衡

按量付费实例节省停机模式

参与者：妍希、巴梨 等 6 人 | 更新时间：2024-04-10 17:48:46

产品详情 相关技术圈 | 我的收藏

通过节省停机模式停止按量付费实例，可以在保留服务器的数据和配置信息的同时，节省部分资源使用成本。开启节省停机模式后，不再收取计算资源（vCPU和内存）、固定公网IP费用，云盘（系统盘和数据盘）、弹性公网IP费用、镜像等资源继续收费。本文介绍节省停机模式的使用限制、计费说明、风险提示，以及如何开启节省停机模式。

使用限制

支持节省停机模式的实例必须同时满足以下条件：

- 网络类型为专有网络
您可以将经典网络下的实例迁移至专有网络，具体操作，请参见[ECS实例从经典网络迁移到专有网络](#)。
- 计费方式为按量付费（包括抢占式实例）
您可以将包年包月实例转换为按量付费实例，具体操作，请参见[包年包月转按量付费](#)。
- 实例规格族不包含本地存储
包含本地存储能力的实例规格族不支持节省停机模式，例如大数据型、本地SSD型等。关于如何查询包含本地存储的实例规格，请参见[实例规格族](#)中的本地存储列。
- 实例规格族不包含持久内存
包含持久内存的实例规格族不支持节省停机模式，例如re6p、re6p-redis。关于如何查询包含持久内存的实例规格，请参见[实例规格族](#)中的持久内存列。

ClickHouse企业版：总结

| 对比维度 | | ClickHouse 开源自建 | ClickHouse 企业版 |
|------|-------|---------------------------|--------------------------------------|
| 成本 | 计算 | 保持峰值资源水位 | 按需动态调整资源，可节约~50% 计算资源 |
| | | 计算和存储绑定，数据容灾需购买两份计算资源 | 计算和存储分离，数据冗余无需冗余计算资源，集群容灾可降低~50%的成本呢 |
| | 存储 | 预估容量，冗余20%-40%空闲资源 | 存储按实际使用量计费，无空闲资源浪费，节约20%-40%存储成本 |
| | | 本地SSD存储，单位成本高 | 共享对象存储，单位价格减少 88% |
| 性能 | 扩展效率 | 全量数据迁移和Rebalance，天/小时级完成 | “零”数据迁移 秒级完成扩展 |
| | 数据更新 | 异步Mutation 更新，无实时性预期 | Lightweight update&delete 实时更新 |
| 稳定性 | 容灾高可用 | 双副本支持，成本翻倍 | 无需副本支持，默认集群分布式高可用 |
| 易用性 | 易用性 | 需要创建分布式表和本地表，使用复杂手动升降配和扩容 | 去除分布式表依赖，更简单 |
| | | | 自动集群扩缩 |

THANKS