阿里云
奥运会官方云服务合作伙伴

PolarDB HTAP 详解

钉钉扫码进群沟通交流

PolarDB HTAP 详解

01 背景

- 挑战
- 业界方案

02 原理

- 架构特性
- 分布式MPP执行引擎
- Serverless弹性扩展
- 消除倾斜

03 功能特性

- Parallel Query
- Parallel DML
- 索引构建加速

04 性能

- 对比单机并行
- 对比传统MPP数据库
- 索引构建加速

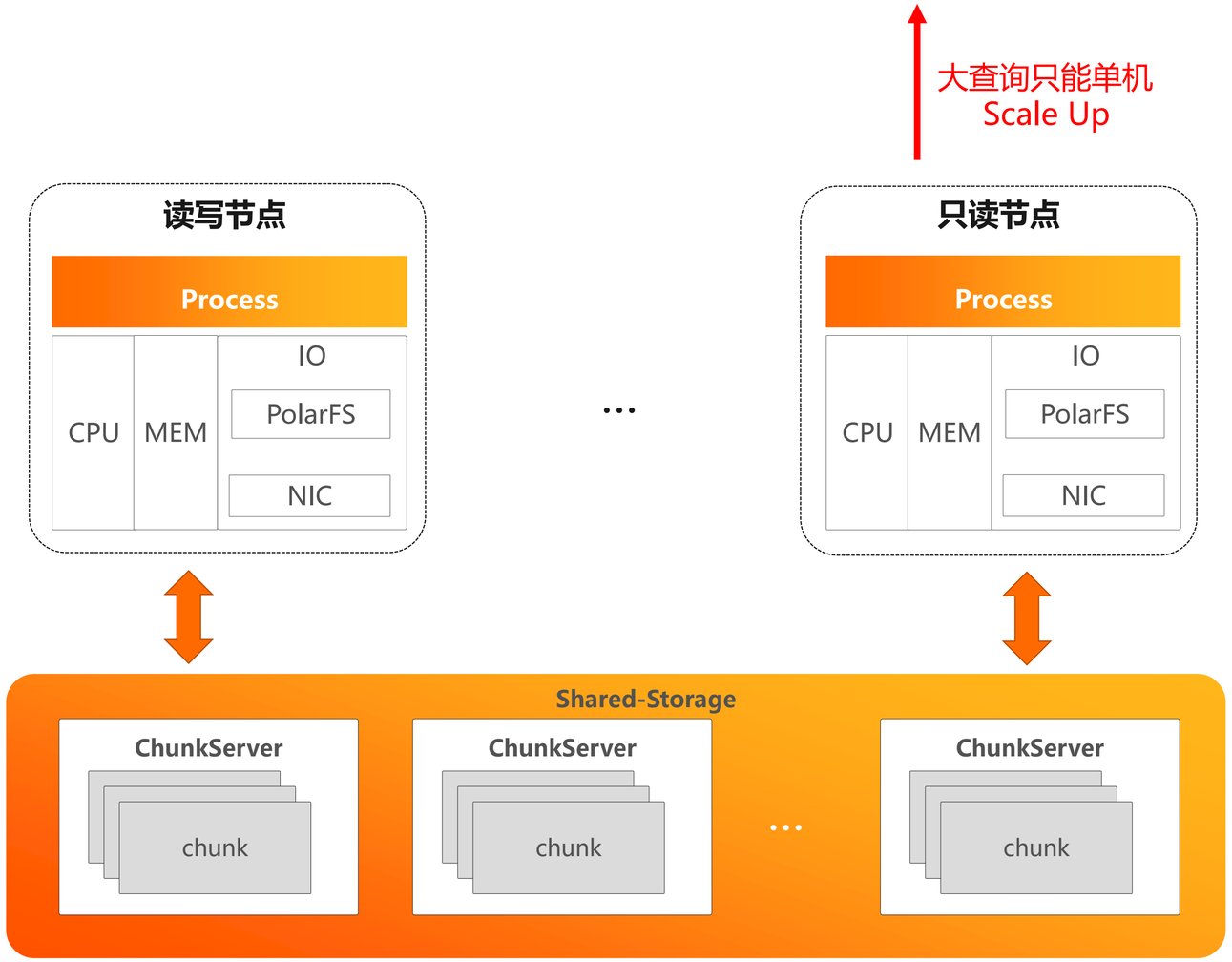
PolarDB HTAP 背景

原生PolarDB PostgreSQL处理AP查询的挑战：

- 无法发挥多个计算节点的CPU/MEM，不能Scale Out
- 无法发挥共享存储池的大吞吐能力

业界HTAP解决方案：

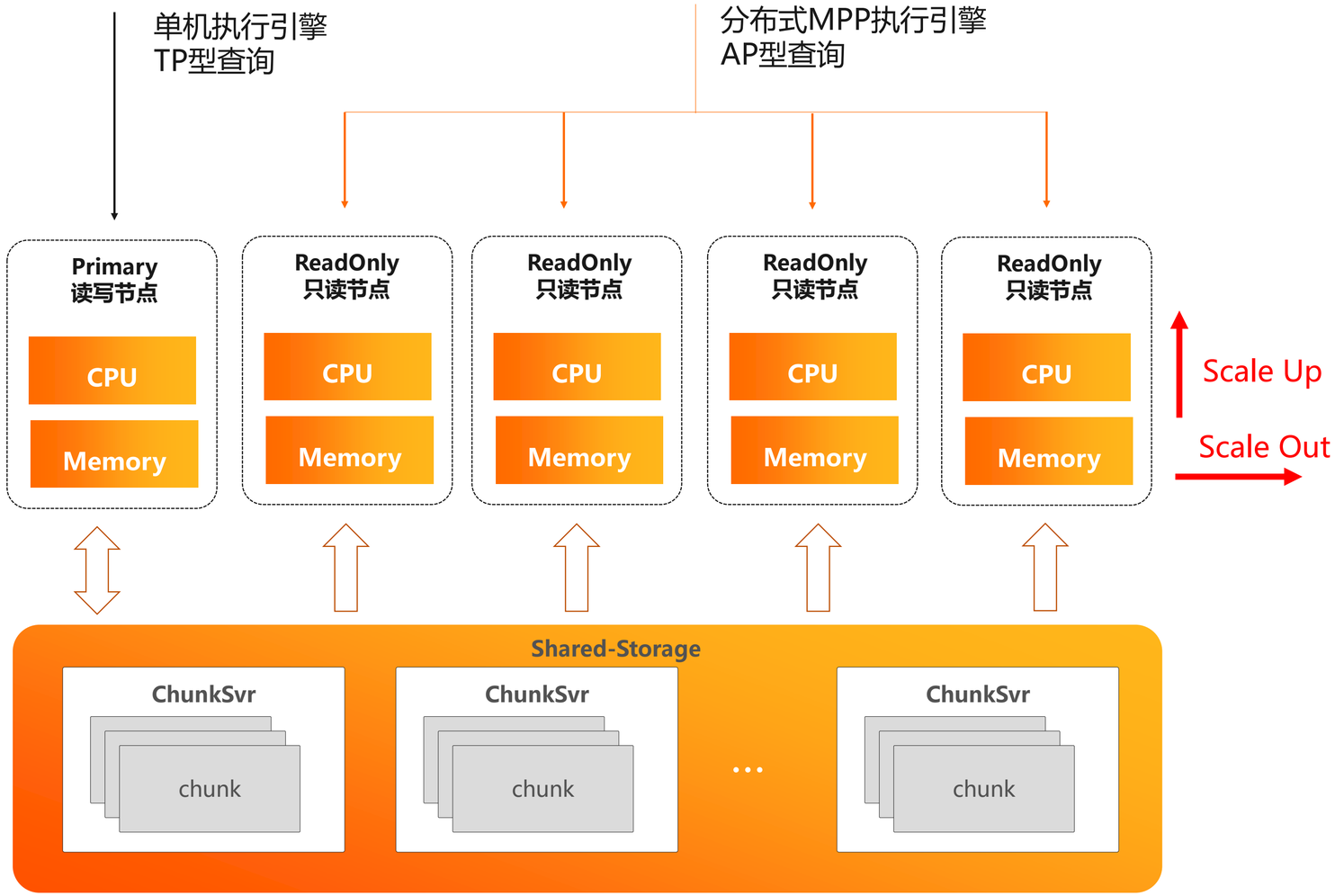
- TP和AP在存储/计算上都分离
  - TP数据导入到AP系统，有延迟，时效性不高
  - 增加冗余AP存储，成本增加
  - 增加AP系统，运维难度增加
- TP和AP在存储/计算上都共享
  - 在执行时AP、TP查询或多或少相互影响
  - 受限TP系统， AP比重增大时，无法快速弹性Scale Out
- TP和AP在存储上共享，在计算上分离



PolarDB HTAP 原理

架构特性

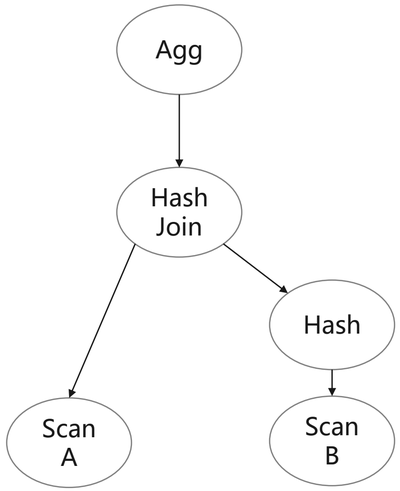
- ✓ **一体化存储** (毫秒级数据新鲜度)
  - TP/AP共享一套存储数据，减少存储成本，提高查询时效
- ✓ **TP/AP物理隔离**（杜绝CPU/MEM相互影响）：
  - 单机执行引擎：在RW/RO节点上, 处理高并发TP查询
  - 分布式MPP执行引擎: 在RO节点, 处理复杂大AP查询
- ✓ **Serverless弹性扩展**
  - 任何一个RO节点均可发起MPP查询
  - ScaleOut：弹性调整MPP执行节点范围
  - ScaleUp：弹性调整MPP单机并行度
- ✓ **消除倾斜**
  - 消除数据倾斜、计算倾斜
  - 充分考虑PG Buffer Pool亲和性



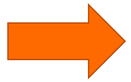
PolarDB HTAP 原理

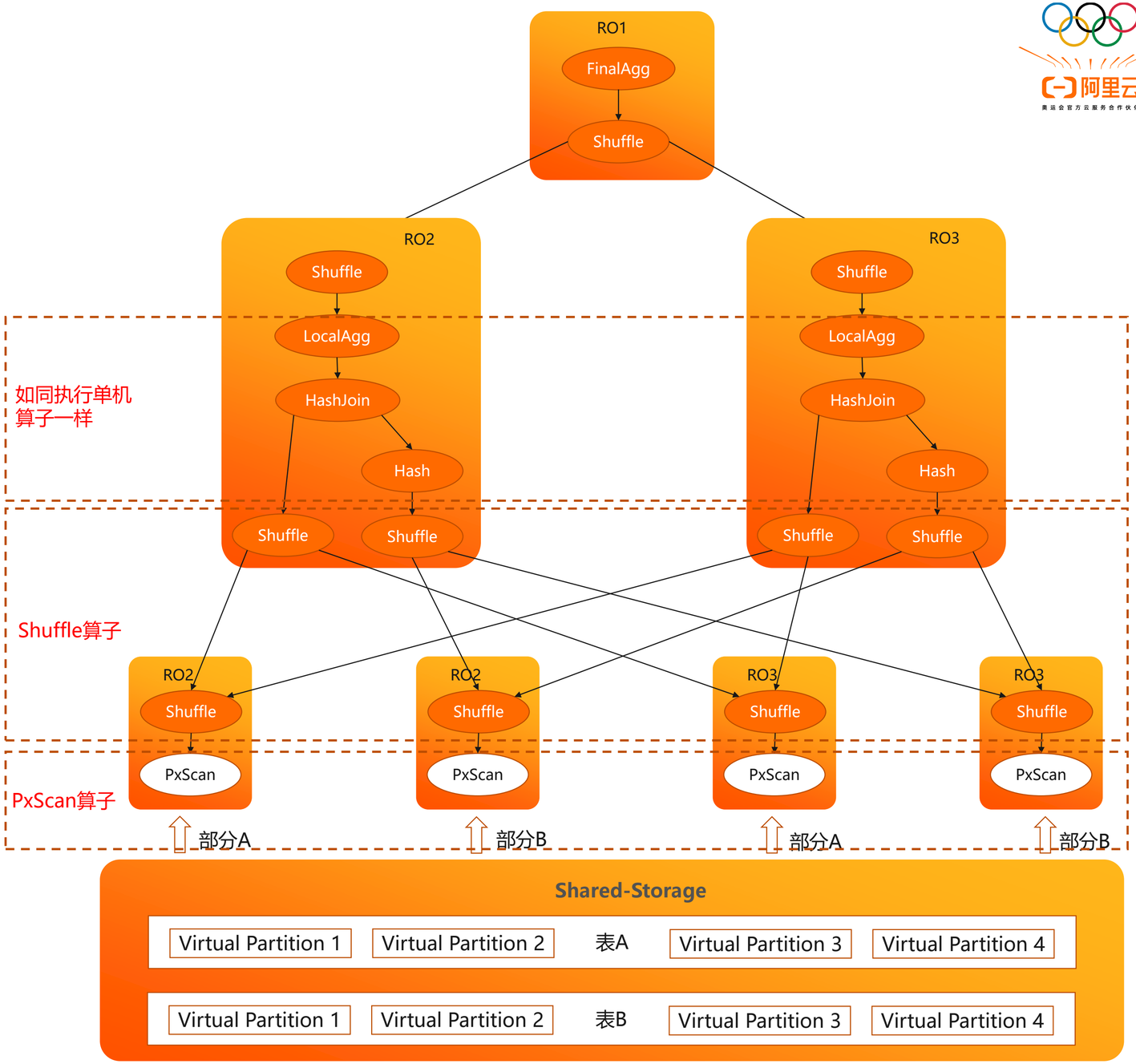
分布式MPP执行引擎

- Shuffle算子屏蔽数据分布
- PxScan算子屏蔽共享存储



《Volcano, an Extensible and Parallel Query Evaluation System》

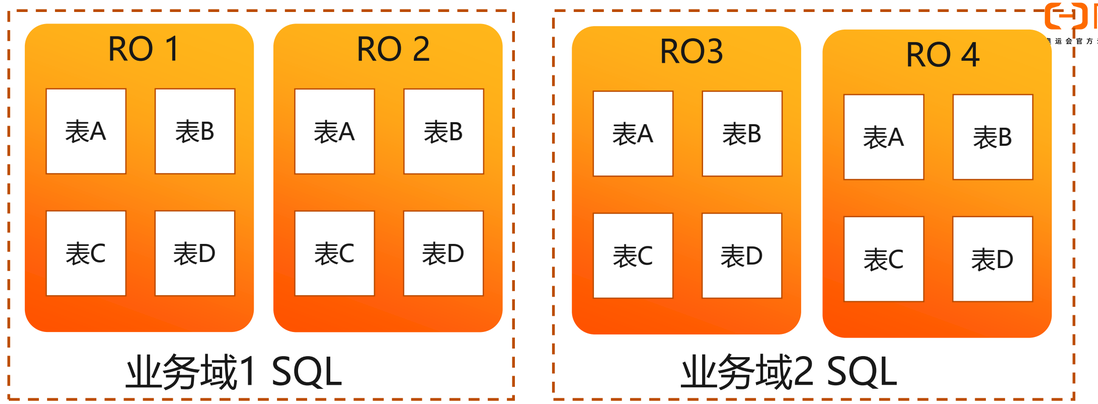


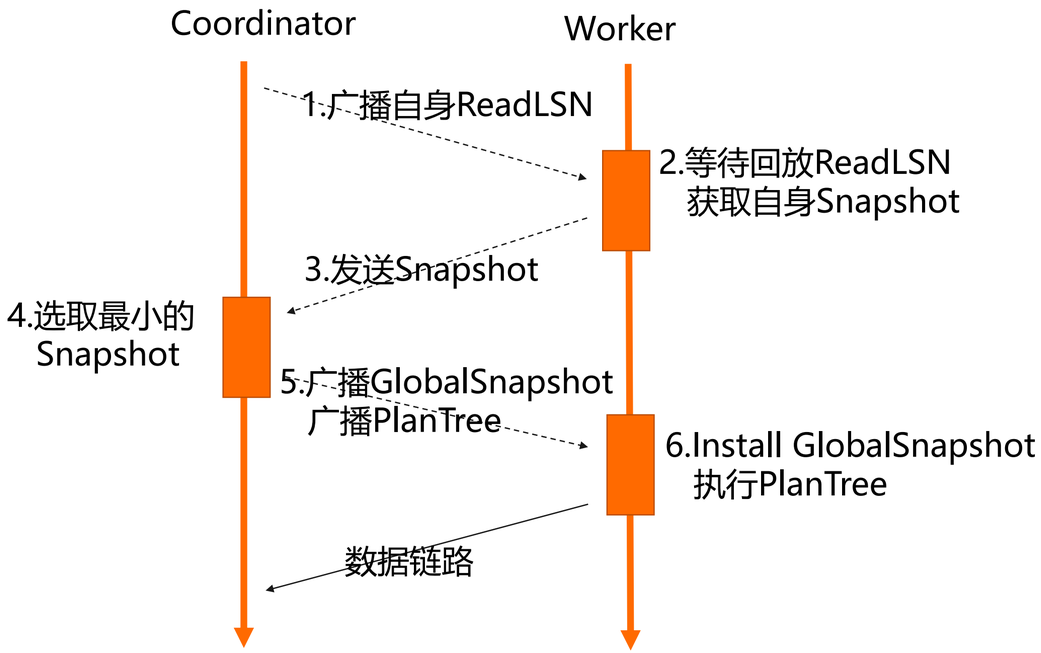


PolarDB HTAP 原理

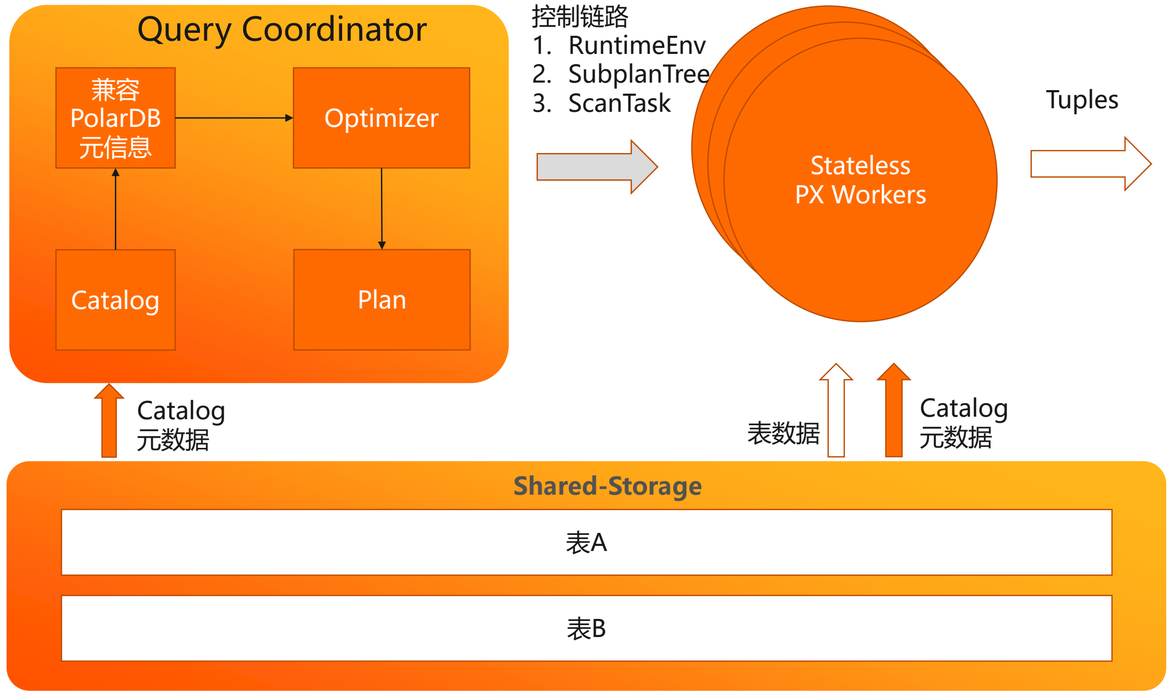
Serverless弹性扩展

- 消除Coordinator单点
- ScaleOut：MPP节点范围弹性扩展
- ScaleUp：单机并行度弹性扩展
- 弹性调度策略





分布式事务一致性保障



Coordinator/Worker全链路无状态

PolarDB HTAP 原理

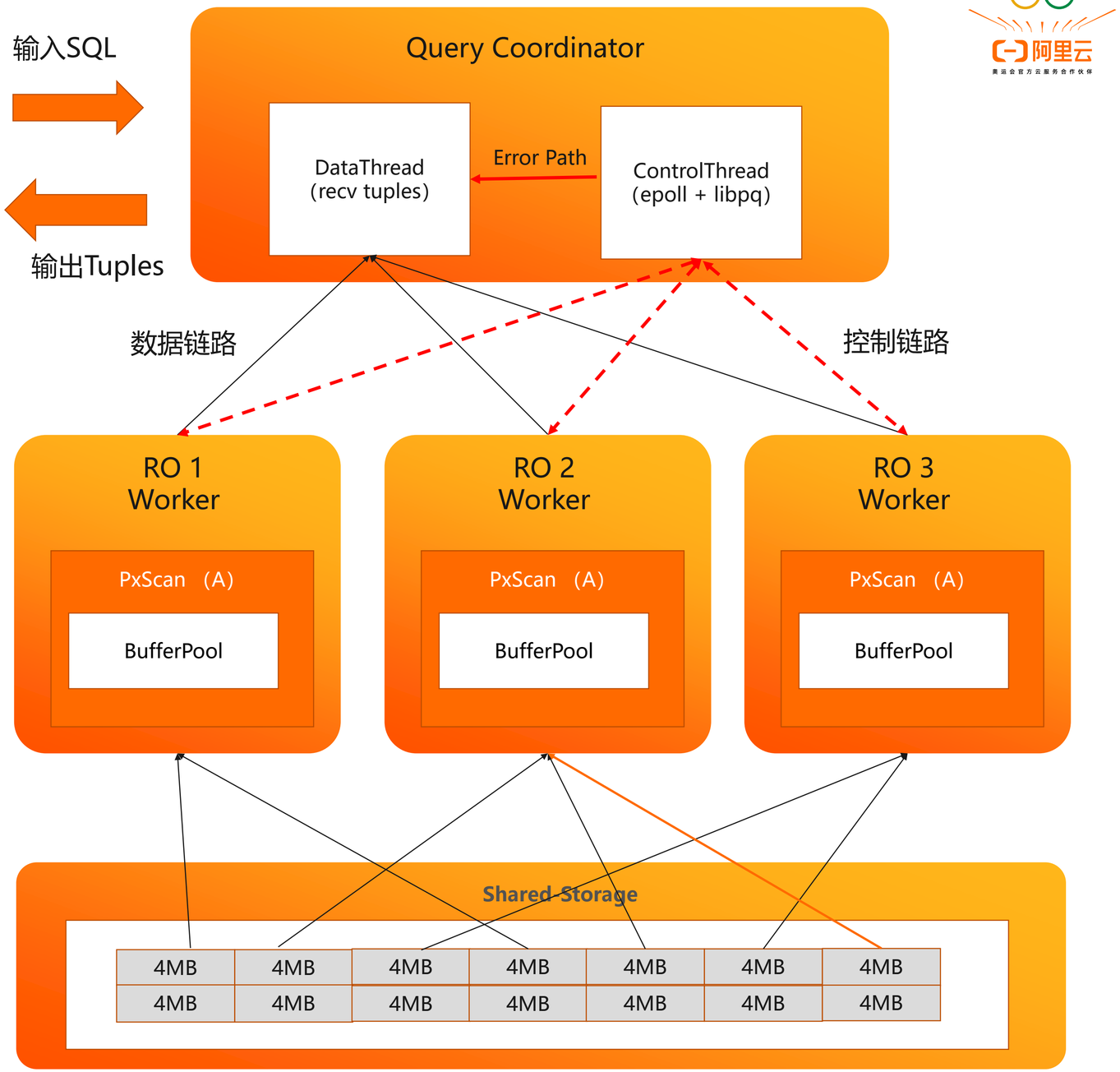
消除倾斜

传统 MPP 固有的问题：

- 数据倾斜：数据打散不均衡、大对象TOAST表存储
- 计算倾斜：不同节点的事务、Buffer、网络、IO抖动，也会导致计算速度不均衡

自适应扫描机制：

- 采用Coordinator节点协调、Worker节点询问的工作模式
- DataThread 负责收集汇总元组
- ControlThread 负责控制每个扫描算子的扫描进度，能者多劳
- 考虑对BufferPool的亲和性， 保证每个Worker尽量扫描固定的数据块



PolarDB HTAP 功能特性

Parallel Query

基础算子全支持	共享存储算子优化	分区表支持	并行度控制	Serverless弹性扩展
SeqScan IndexScan IndexOnlyScan BitmapScan SubqueryScan HashJoin MergeJoin NestloopJoin Agg WindowAgg Material Union All Append CTE PLSQL …	Shuffle ShareSeqScan ShareIndexScan LeftOuterIndex NestloopJoin LeftAntiSemiJoin NotIn …	Hash/Range/List 多级分区 分区静态裁剪 分区动态裁剪 Partition Wise Join …	全局级别并行度 表级别并行度 会话级别并行度 查询级别并行度 …	任意节点发起MPP MPP节点范围任意 组合 集群拓扑信息 自动维护 支持共享存储模式 支持主备库模式 支持三节点模式 …

PolarDB HTAP 功能特性

Parallel DML

一写多读	多写多读	一写多读	多写多读
非分区表	非分区表	分区表	分区表
RW上单Workers写 RO上多Workers读 支持Insert Into Select 支持Update 支持Delete	RW上多Workers写 RO上多Workers读 支持Insert Into Select 支持Update 支持Delete	RW上单Workers写 RO上多Workers读 支持Insert Into Select	RW上多Workers写 RO上多Workers读 支持Insert Into Select

PolarDB HTAP 功能特性

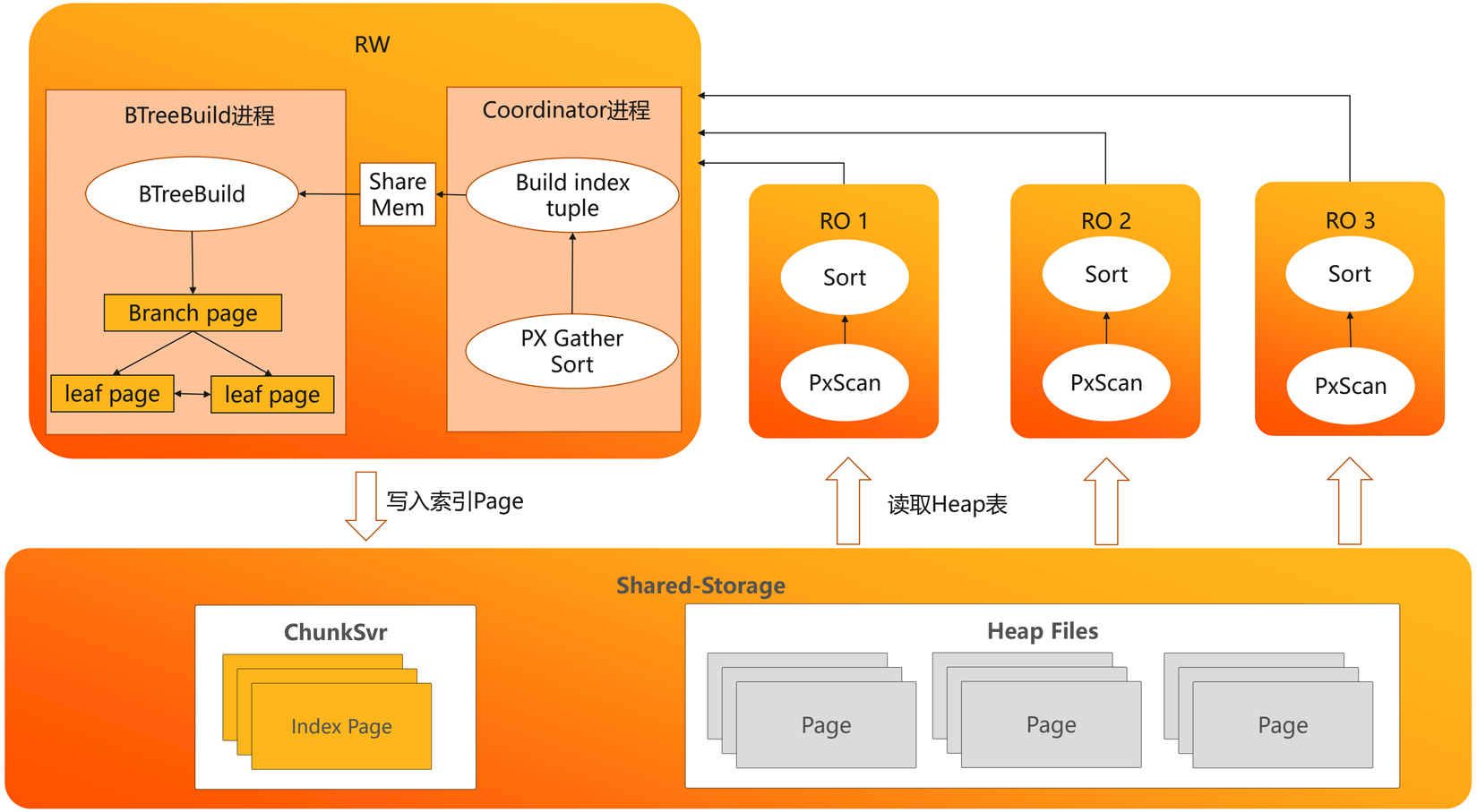
索引构建加速

优化点:

- 分布式MPP并行来加速排序过程
- 全流水化
- 批量写入

支持范围:

- Create BTree Index
- Create BTree Index Concurrently



PolarDB HTAP 性能

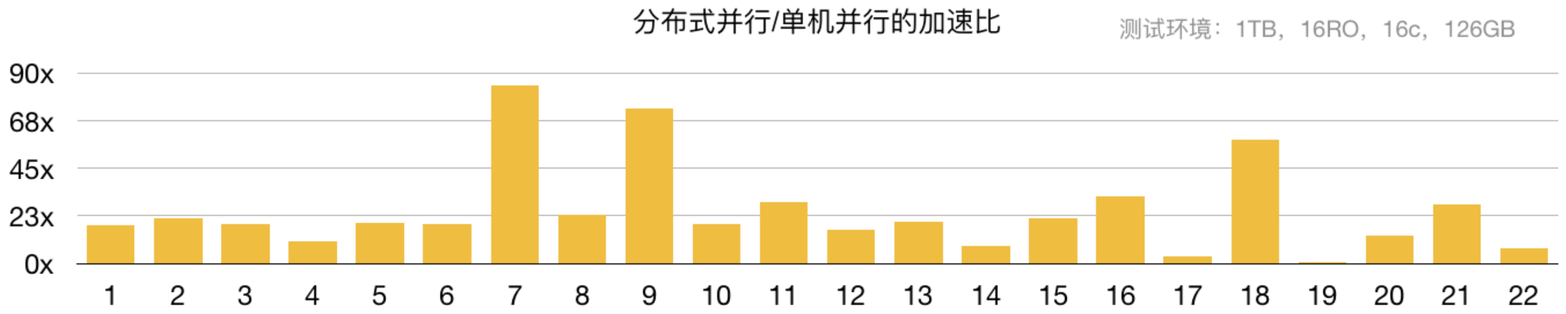
TPCH 单机并行 vs 分布式MPP

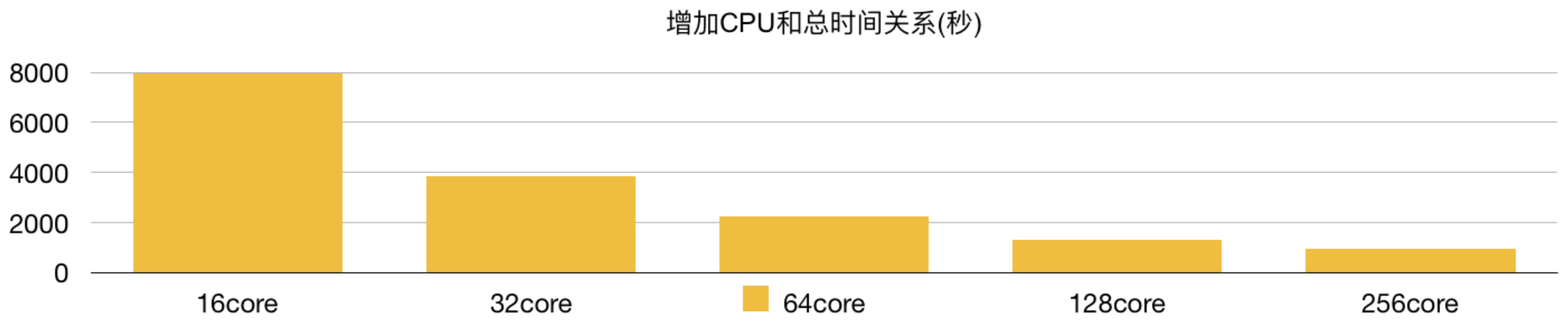
分布式MPP并行/单机并行的加速比:

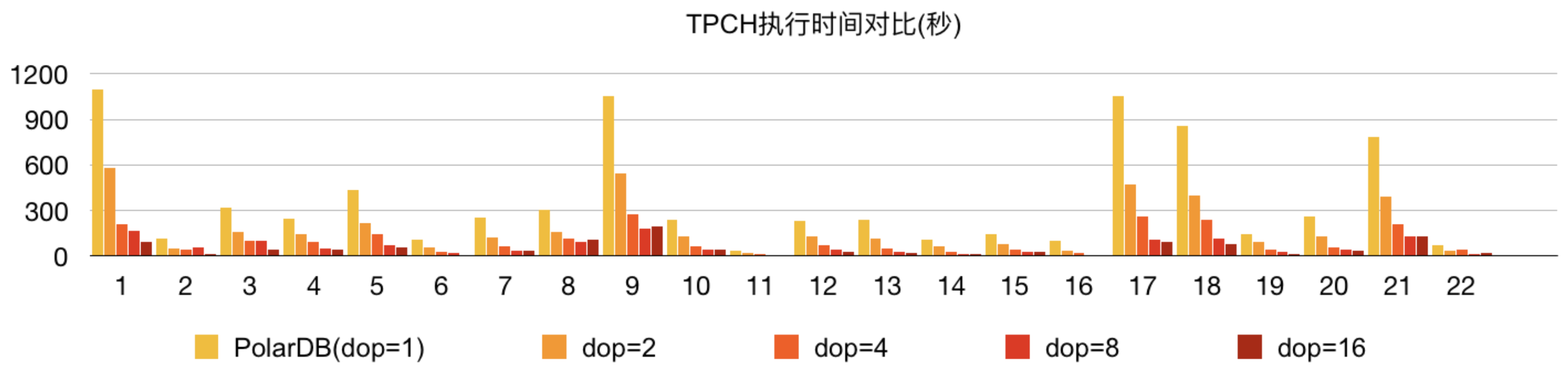
- 3个SQL加速比60x
- 19个SQL加速比10x
- 平均加速23倍

增加总CPU和执行时间关系:

- 16到128core线性提升
- 单条SQL线性提升







PolarDB HTAP 性能

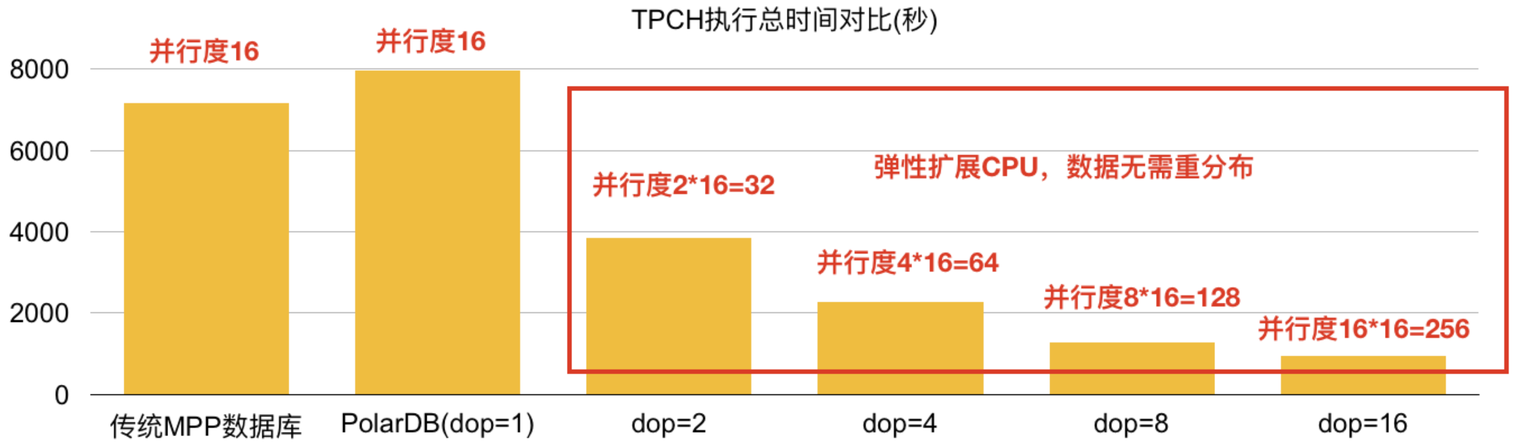
TPCH vs 传统MPP数据库

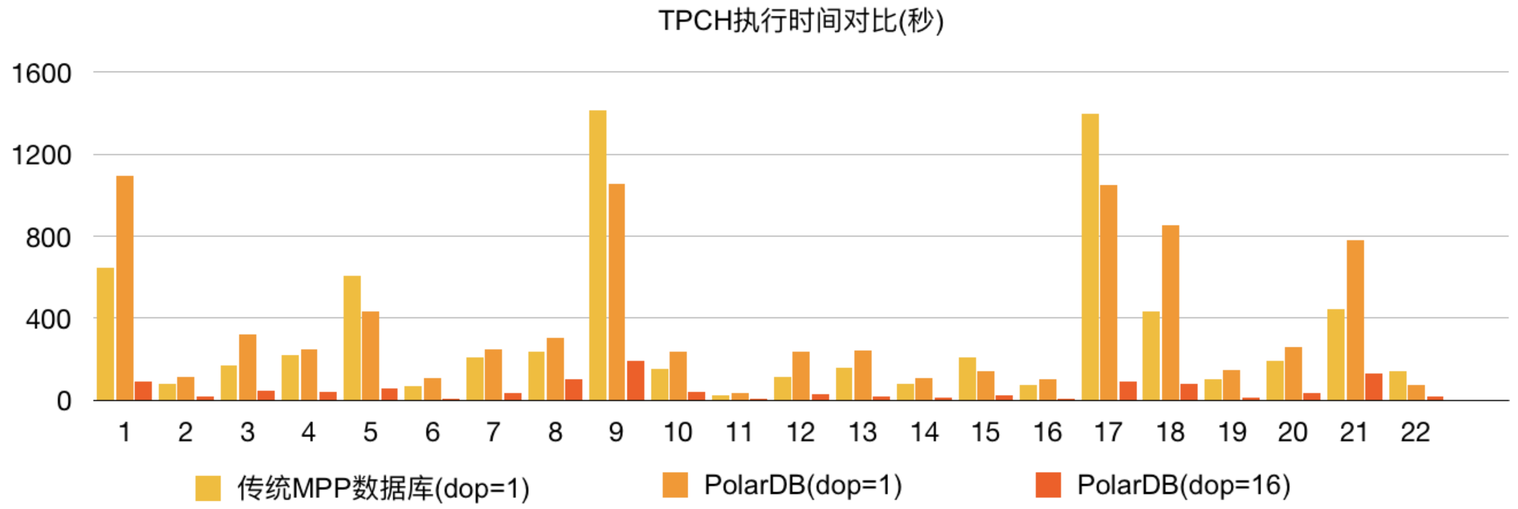
相同并行度时（DOP=1）：

- PolarDB性能是传统MPP数据库的90%
  - 数据分布不同导致

PolarDB弹性扩展到（ DOP =8）时：

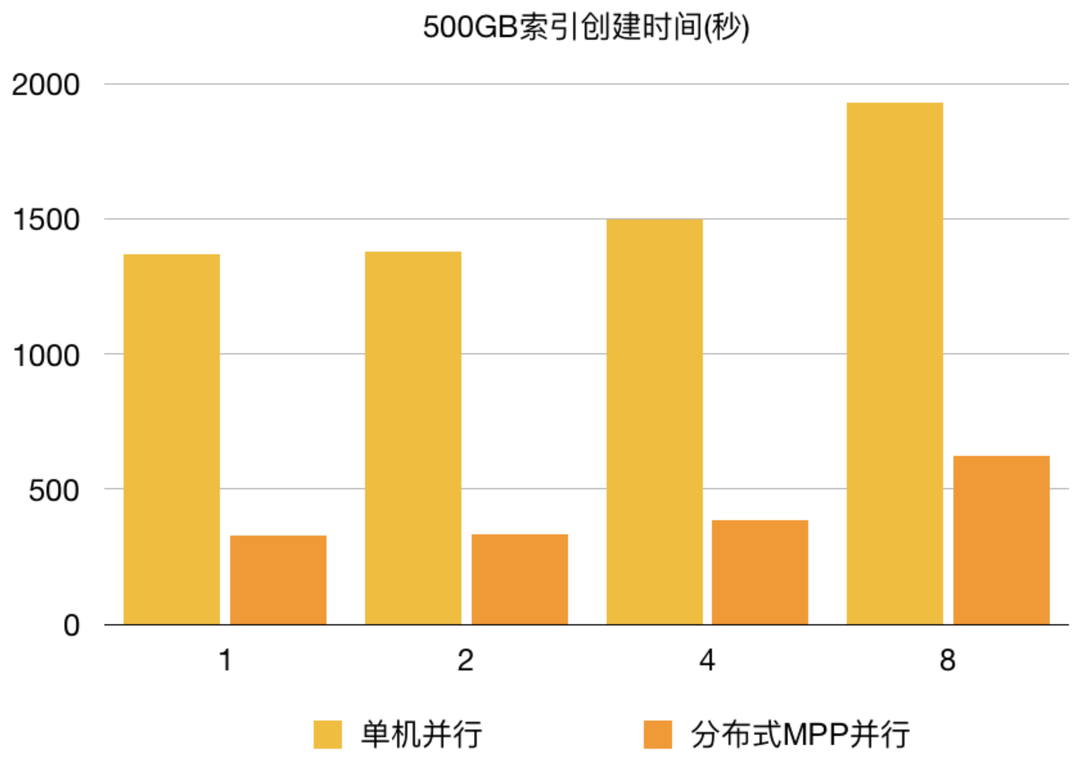
- PolarDB性能是传统MPP数据库的5.6倍





PolarDB HTAP 性能

加速索引构建，性能提升4~5倍



观看回放：
https://developer.aliyun.com/topic/PolarDB_release

钉钉扫描下方二维码，获取更多
新品发布会资料，直播信息。
或搜索钉钉群：30675445

开源PolarDB企业级架构重磅发布