PolarDB-X: An Elastic Distributed Relational Database For Cloud-native Applications

YunTzu Chen(ychen2334)

Challenges:

- 1. Performance
- 2. Scalability
- 3. One Size Doesnot Fit All

Solution: PolarDB-X

- Cross-DC Transactions.
  - uses HLC-SI(Hybrid Logical Clocks) to achieve data consistency.
- Elasticity
  - separation of computation and storage architecture
- HTAP (Hybrid transactional/analytical processing)
  - optimizer can identify whether a query belongs to TP or AP workload

System Overview

- **Architecture**
  - GMS(Global Meta Service)
  - Load Balancer
  - Computation Node
  - Database Node(PolarDB)
  - Storage Node(PolarFS)

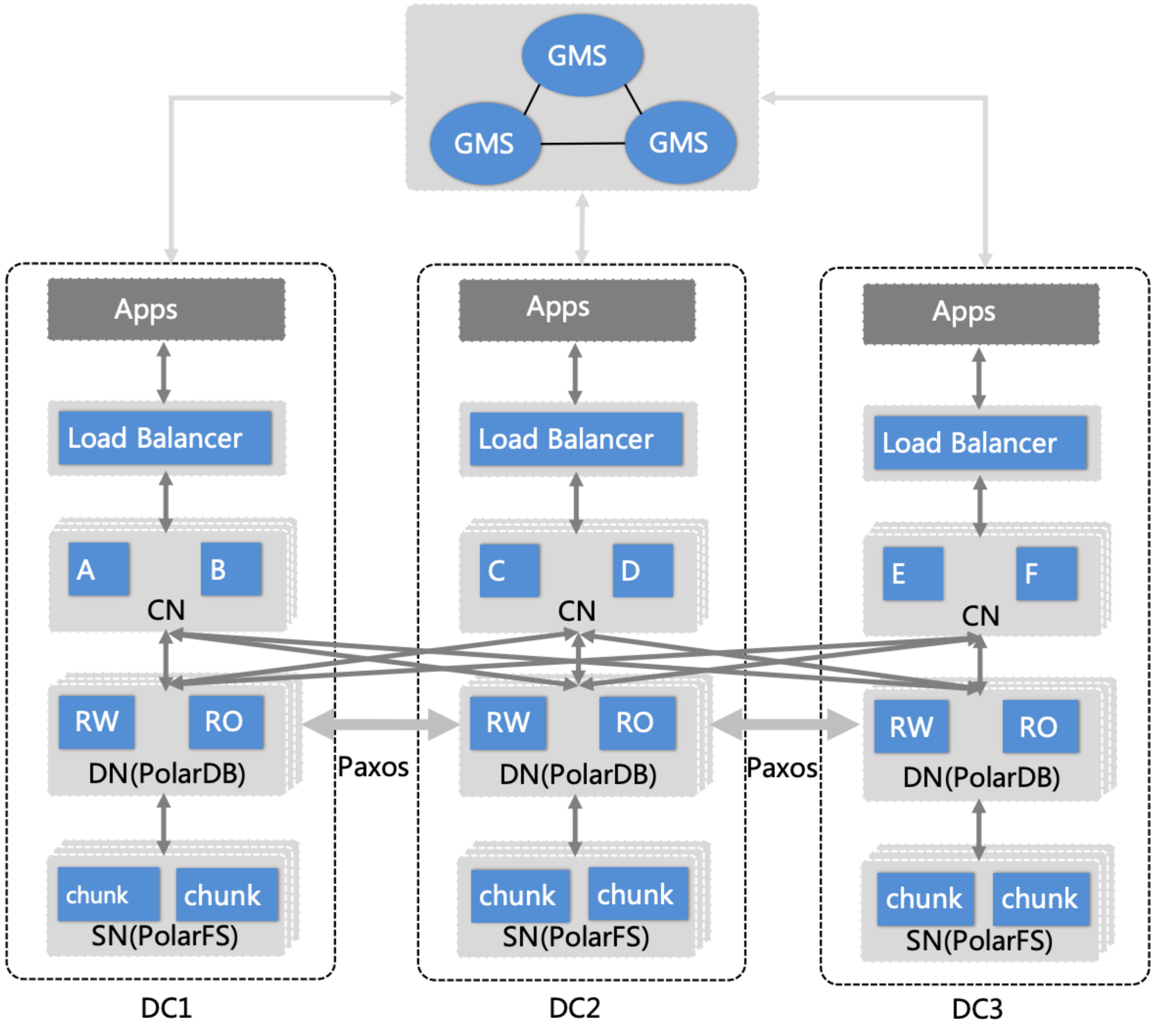


Fig. 2. System architecture of *PolarDB-X*

Transactions

- **HLC-SI**(Hybrid Logical Clocks with Snapshot Isolation)
    - freserved：2；pt：46；lc：16g
    - the logic clock part is not incremented in ClockUpdate and ClockNow
    - minimize the calls to ClockUpdate
    - still preserves the properties of snapshot isolation

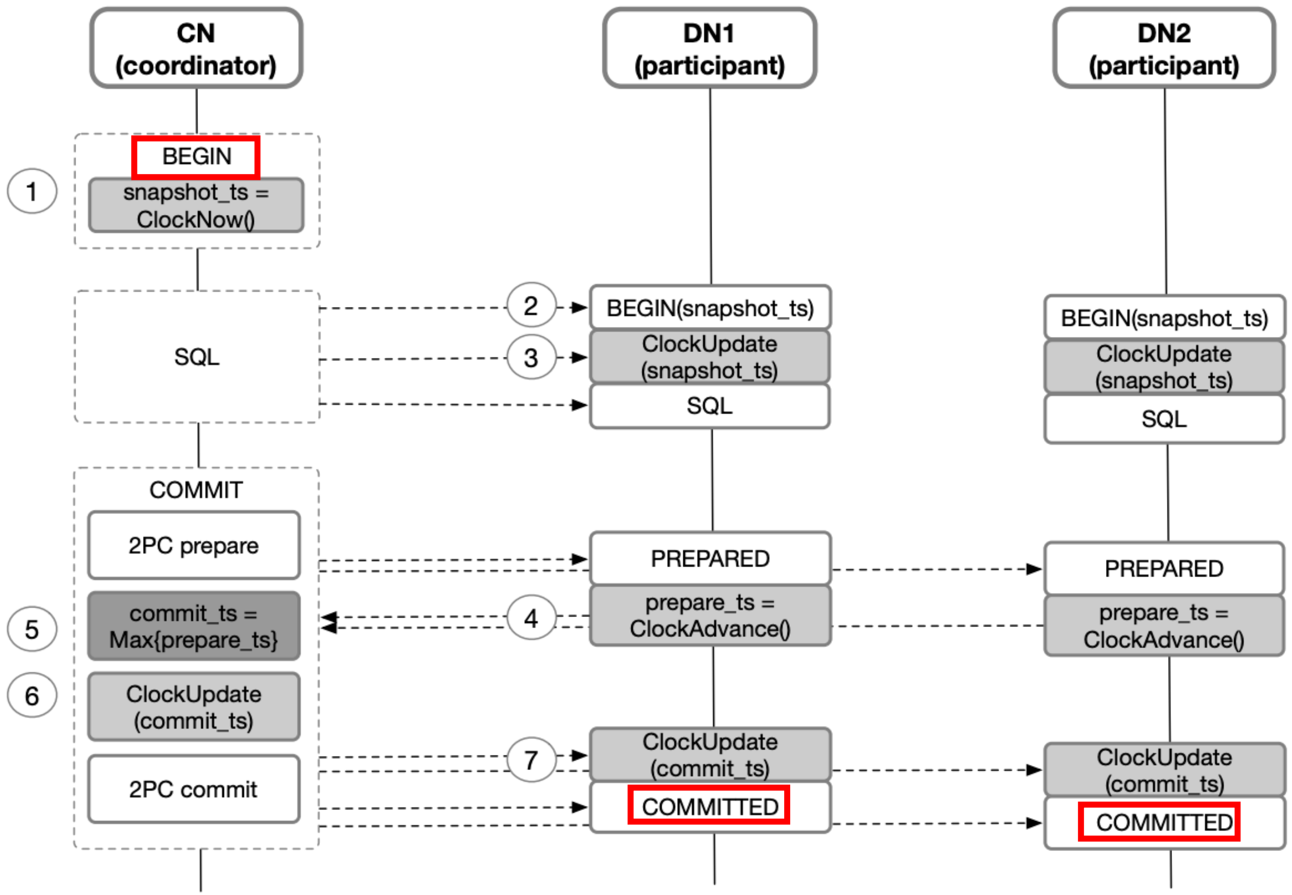


Fig. 4. HLC-SI and two-phase commit

Elasticity

- Multi-Tenancy
  - Allows use of multiple RW nodes for scalable writes
- Design of PolarDB-MT
  - share a global data dictionary and the master RW node manages the data dictionary
- Scale PolarDB-X cluster

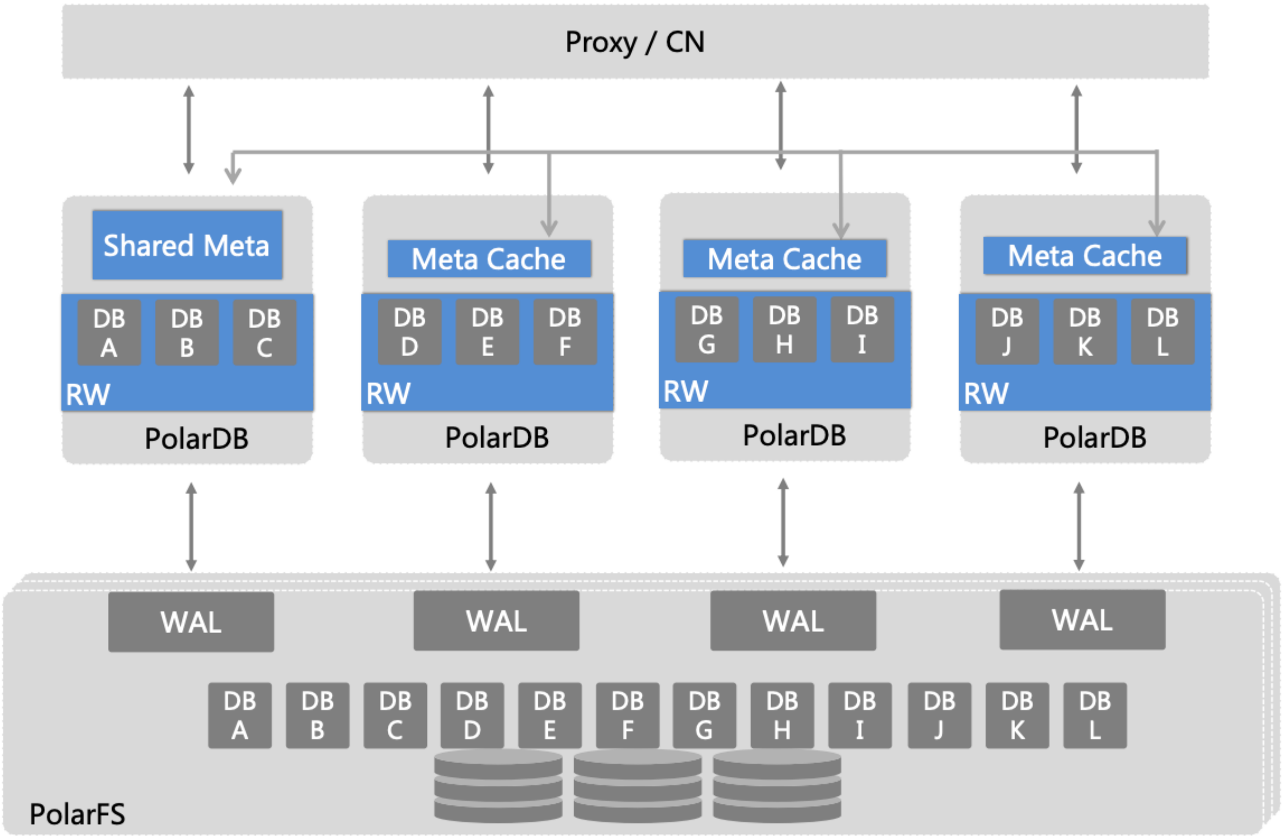


Fig. 5. PolarDB Multi-Tenant Architecture

HTAP

HYBRID TRANSACTIONAL/ANALYTICAL PROCESSING

- **HTAP Optimizer**
  - Request classification and routing
    - equipped with query classification
  - Advantages
    - OLTP workloads will not be delayed from log replication
    - OLAP workloads can obtain good scalability
- **Resource isolation**
  - CPU resource
    - TP Group <----- TP Core Pool
    - AP Group(strictly controlled by cgroups) <----- AP Core Pool, and Slow Query AP Core Pool

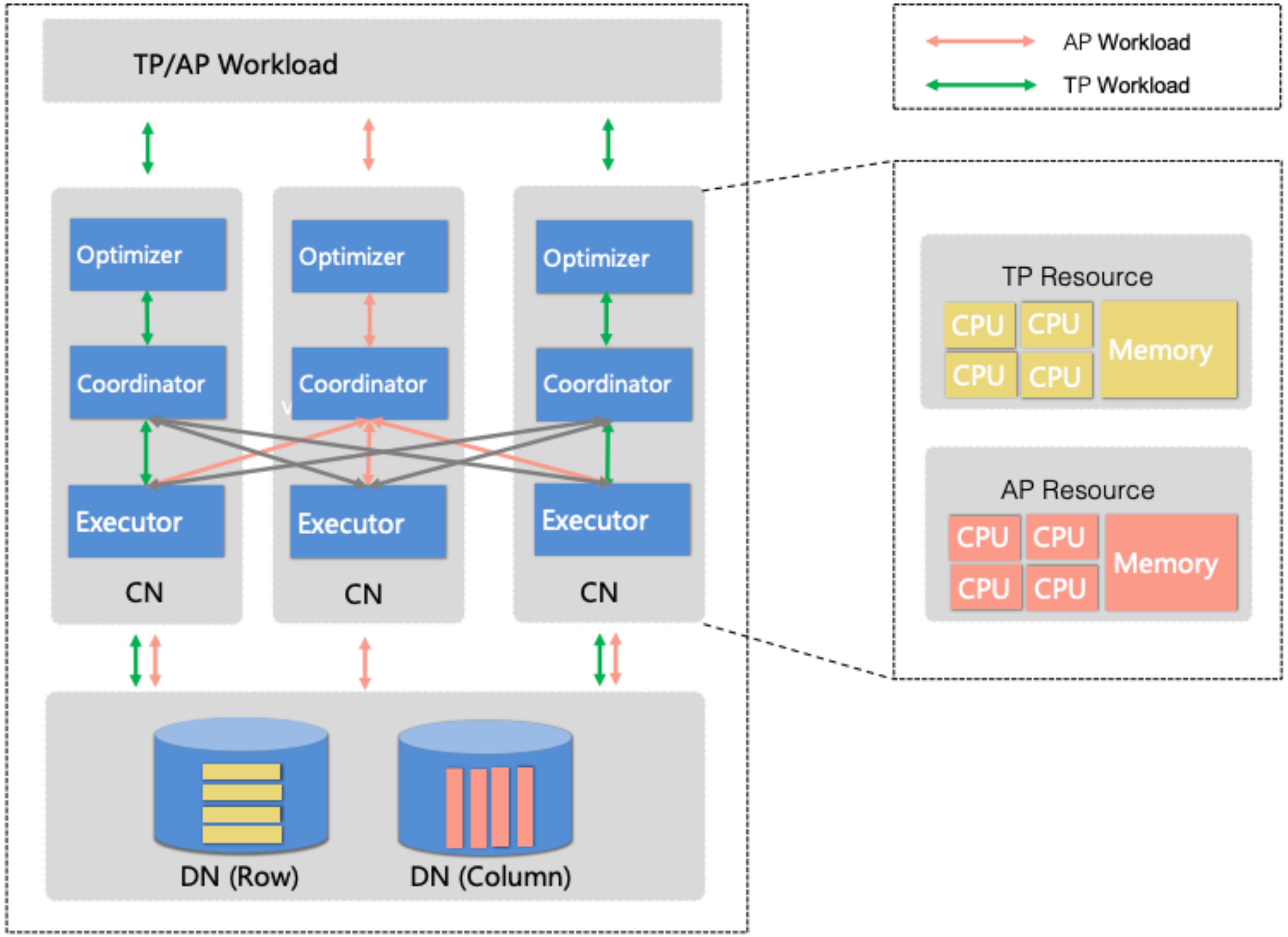
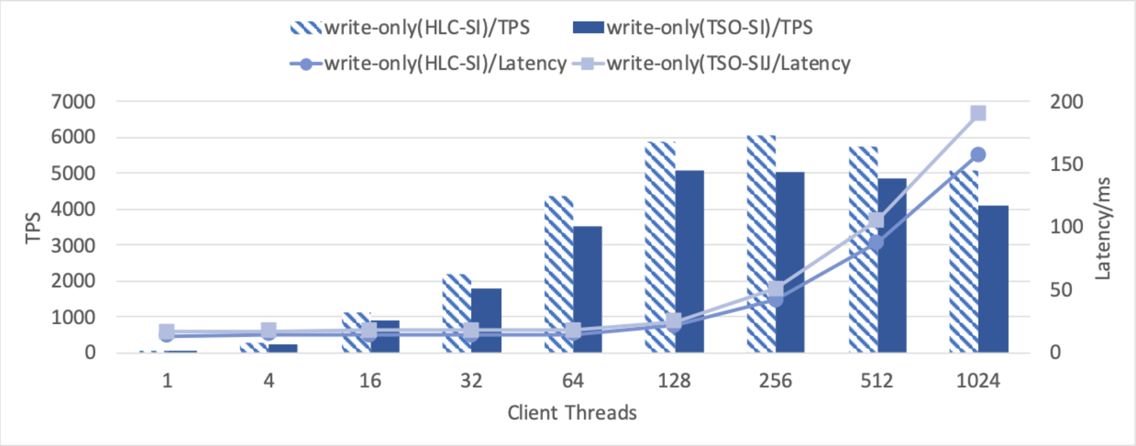


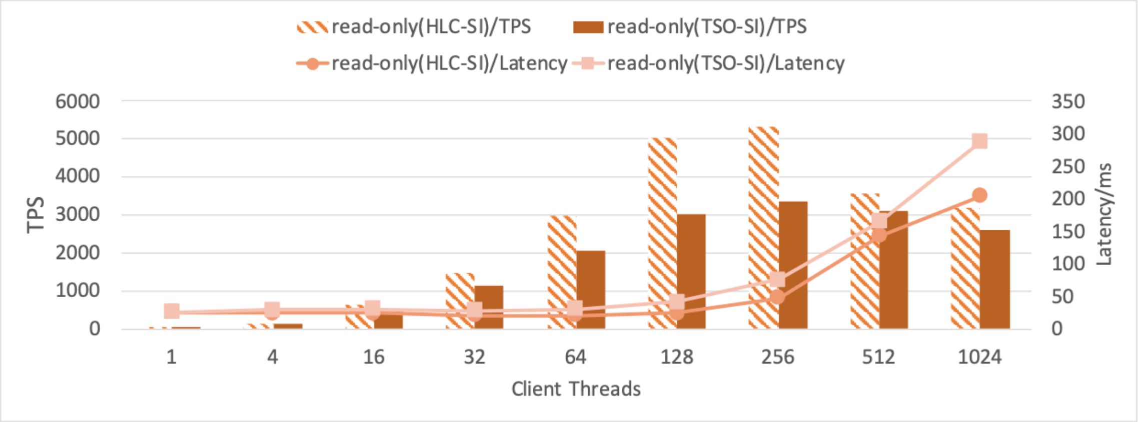
Fig. 6. *PolarDB-X* HTAP framework

Evaluation

- Transactions
  - average transaction latency:
    - TSO-SI > HLC-SI
  - peak write throughput:
    - TSO-SI < HLC-SI



(a) Sysbench Write-Only Transactions

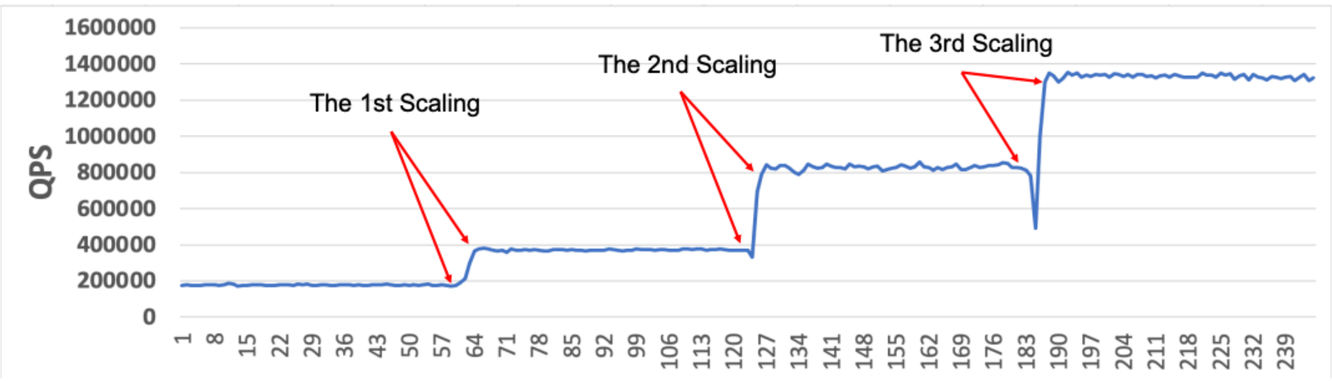


(b) Sysbench Read-Only Transactions

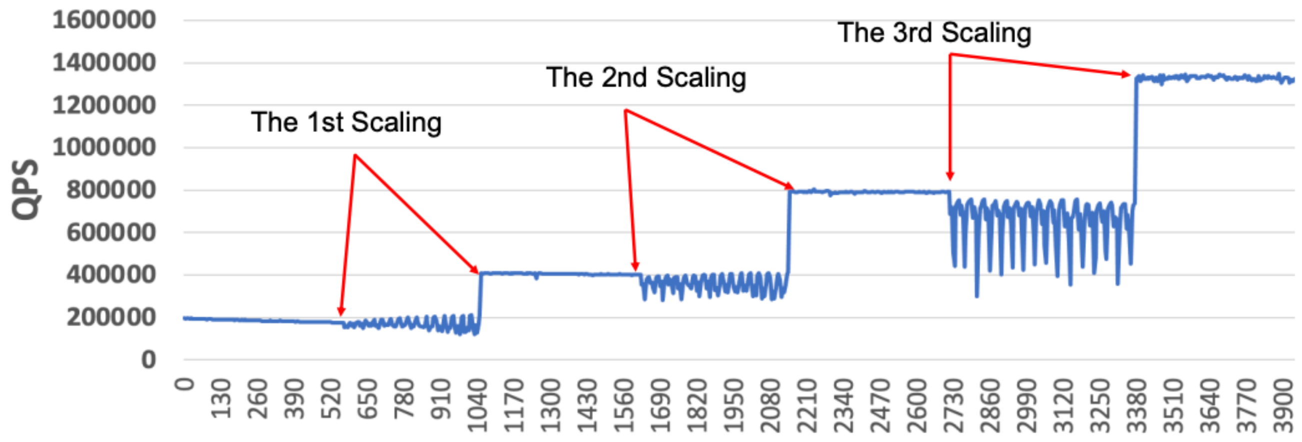
Fig. 7. Comparison of TSO-SI and HLC-SI with Sysbench when deployed across DC

Evaluation

- Elasticity
  - (a) three scaling operations are completed in 4.2, 4.5 and 4.6 seconds
  - (b) it takes 489, 527 and 660 seconds to scale the cluster using data transfer



(a) Scale Using Fast Tenant Migration Mechanism Of PolarDB-MT

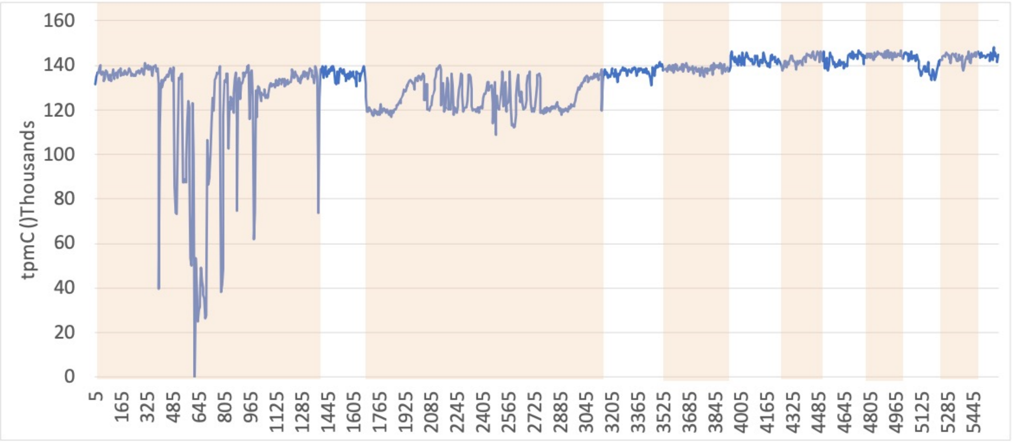


(b) Scale By Copying Data Between Source And Target

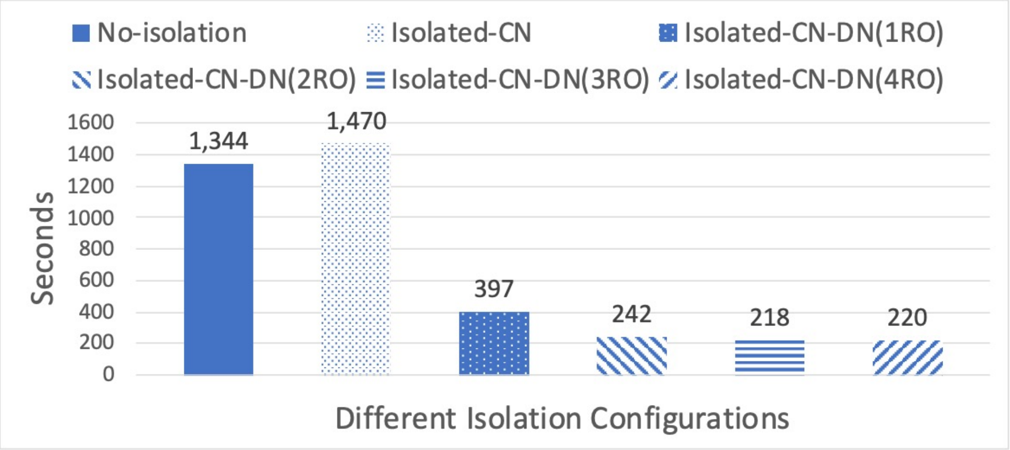
Fig. 8. Comparison of Scaling PolarDB-X using different approaches

Evaluation

- HTAP - Resource Isolation and Scalable RO
    1. the resource isolation switch of CN is turned off, TPC-H is sent to the RW node
    2. the resource isolation switch of CN is turned on, TPC-H is sent to the RW node
    3. last four configurations, use one to four dedicated RO nodes respectively, TPC-H is sent to the RO node



(a) Performance variation of TPC-C while TPC-H runs six times



(b) Latency of each run of TPC-H

Fig. 9. TPC-C and TPC-H Performance when running mixed workloads with different resource isolation and available resources configurations