



AliSQL: Breakthrough for the Future

Wei Hu

Database Kernel Developer @ Alibaba Group

Percona live 2017

Apr 27, 2017

About The Talk



01 / A Bit of History about AliSQL

Introduction of AliSQL and what we did in the past years

02 / X-KV

AliSQL key-value interface

03 / X-Cluster

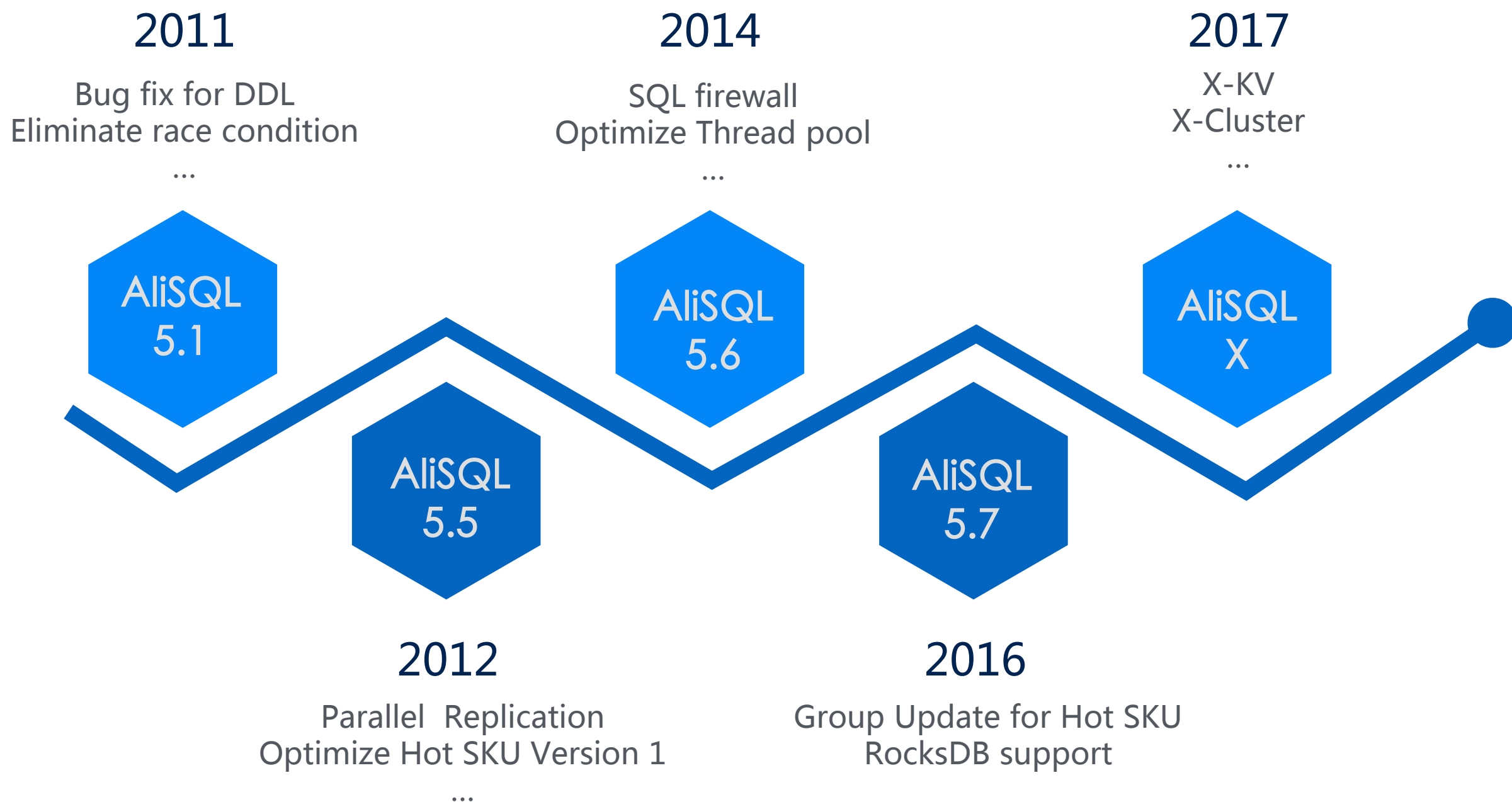
AliSQL cluster solution



HISTORY

A bit of history about AliSQL

Roadmap



First View of AliSQL

Percona live 2016 [What's new in AliSQL- Alibaba's branch of MySQL](#)

commit_on_success hints

Group prepare

Enhanced Thread Pool

Select from Update

Hot SKU Optimization

Dual redo log buffers

Table-level Parallel Replication

PK access optimization

Table Lock optimization for select

Enhanced Semi-Sync Replication

Single row update reached 100,000 per second
By one AliSQL instance last year

Single's Day is Always a challenge

We Need...

- Extreme performance
- Strong data consistency
- Low Cost





X-KV

Key-value interface of AliSQL

What is X-KV

AliSQL KV Interface

An extended Key-Value API for AliSQL which based on InnoDB memcache plugin

- There were so many simple queries according to unique keys
- Data Consistency is very important for us
- SQL parsing and Query Optimization is too expensive

Differences between InnoDB memcache plugin and X-KV ?

What is X-KV

AliSQL KV Interface

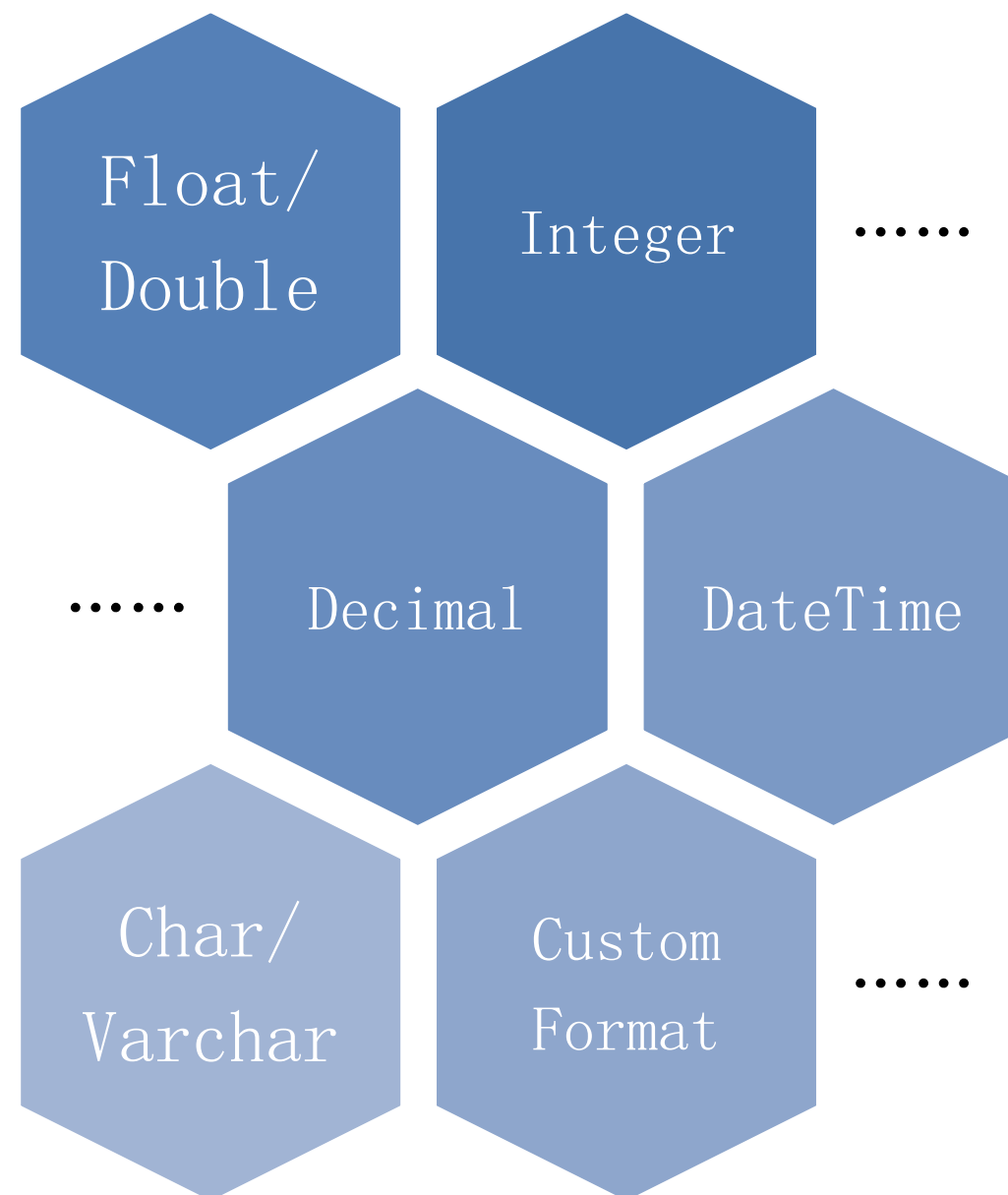
Current InnoDB memcache plugin cannot satisfy our needs

- Support More data types and index types
- Design new protocol
- Optimize the performance
- Easy operation and maintenance

Data Types Support

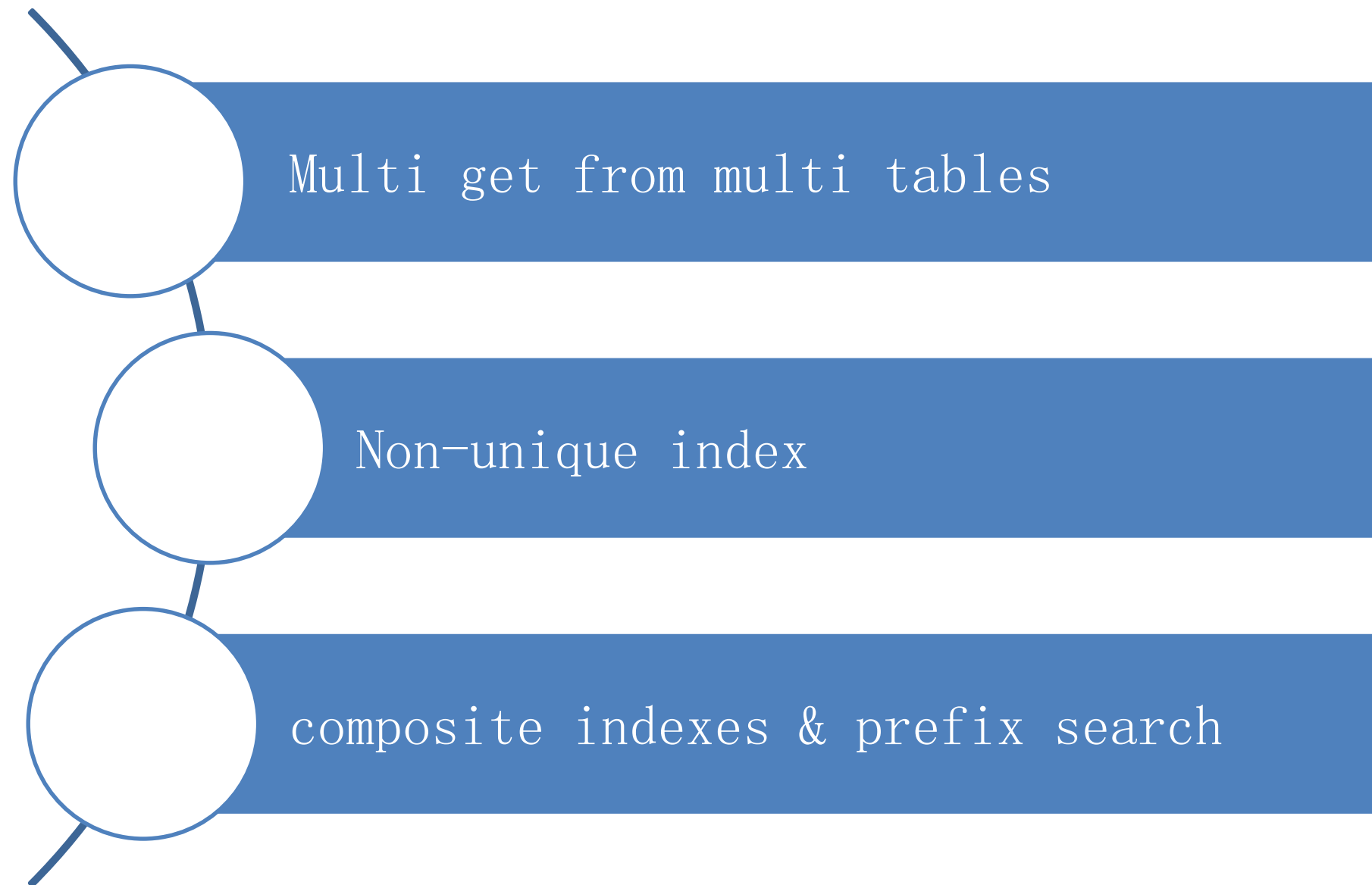
X-KV

Char/Varchar/Blob are not enough!!!



Features extension

X-KV



New Protocol

X-KV

- How to deal with NULL value ?
- How to choose delimiter?
- How to support composite index?

C1 (PK)	C2	C3	C4	C5
1	NULL	S ad	Steve#	3000

- Field = Meta info + Data
- Comosit index support :
get [@@name.]key1 [key2] ... [keyN]
- Meta info format:
Version / Length1 / Length2 / Length3 //

Convenience

X-KV

How to deal the cases when user table schema changed

Original solution is to stop request and then reinstall plugin

Unacceptable

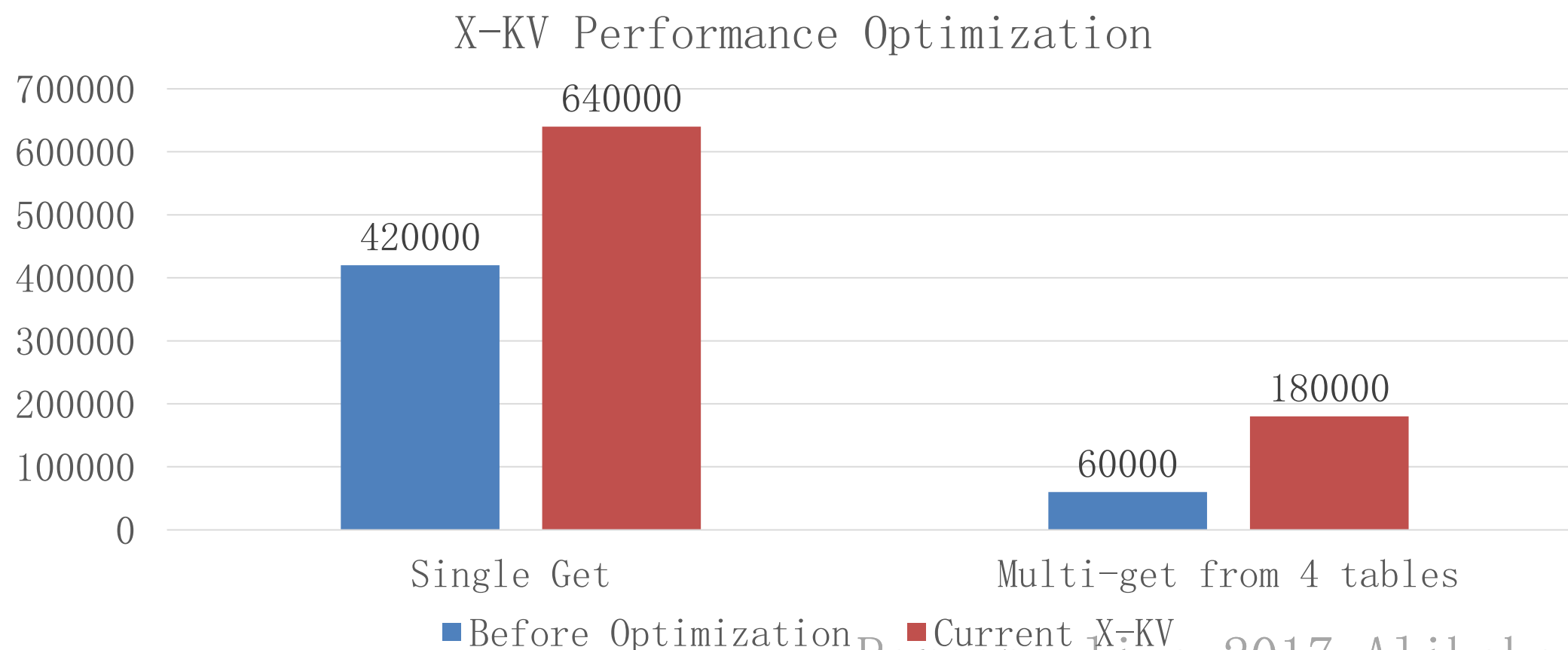
X-KV Support Online Schema Change

- Not block user queries from KV interface
- Configurations are loaded dynamically

X-KV performance optimization

Profiling result shows bottlenecks In our production environment

- Efficiency of serialization
- Allocation of transactions
- Container info verifying





X-Cluster

AliSQL Cluster base on X-Paxos

Background

*Article: Existential Consistency:
Measuring and Understanding Consistency at Facebook*

Large web services faces a trade-off between stronger forms of consistency and higher performance properties.

For Alibaba:

- Data consistency is the most important thing
- High performance is also indispensable

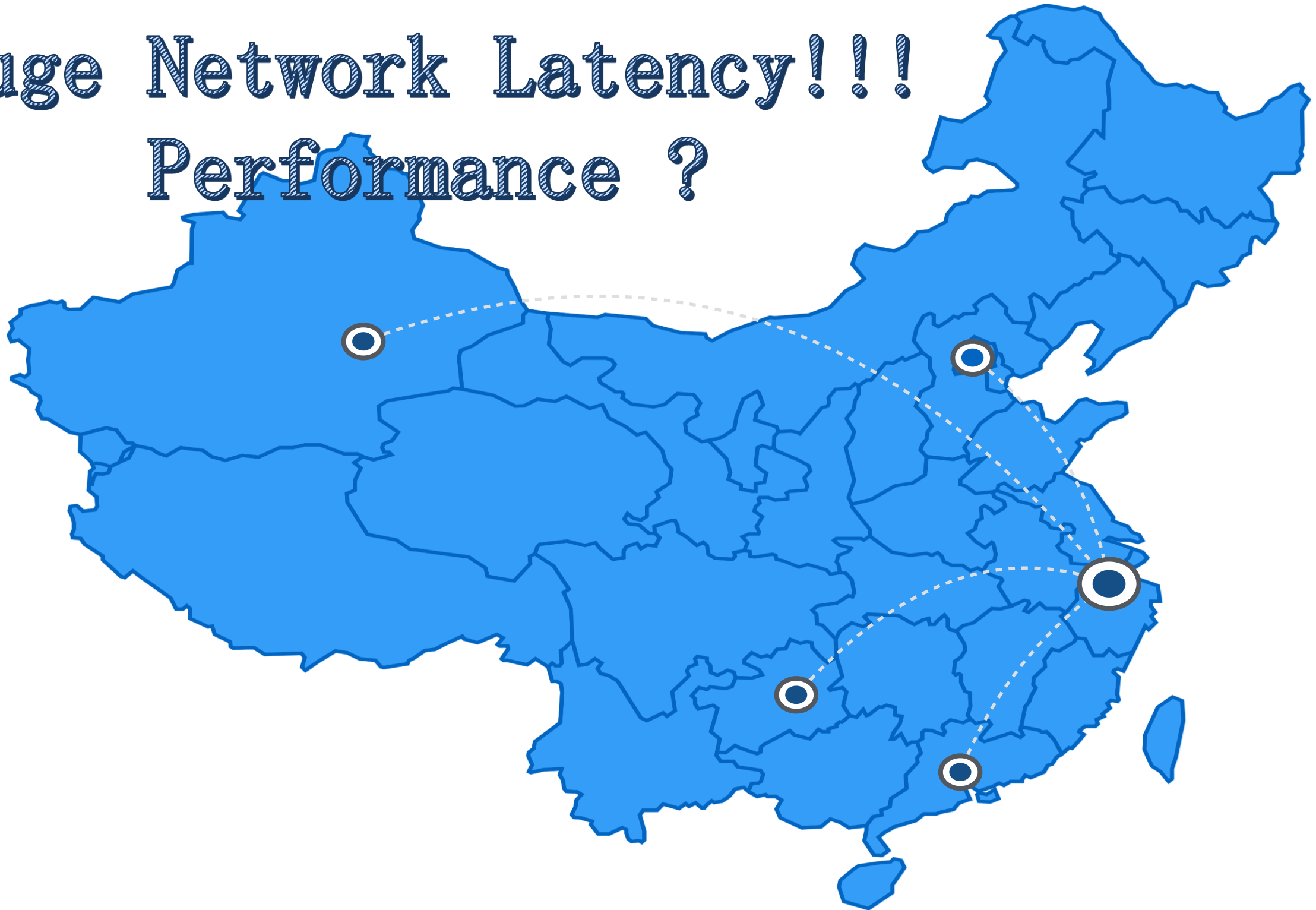
Ideas

The cluster solution must...

- Using Consensus protocol to handle the data consistency problem
- Take advantages of AliSQL
- Using Batching/Pipelining/Asynchronization to boost performance

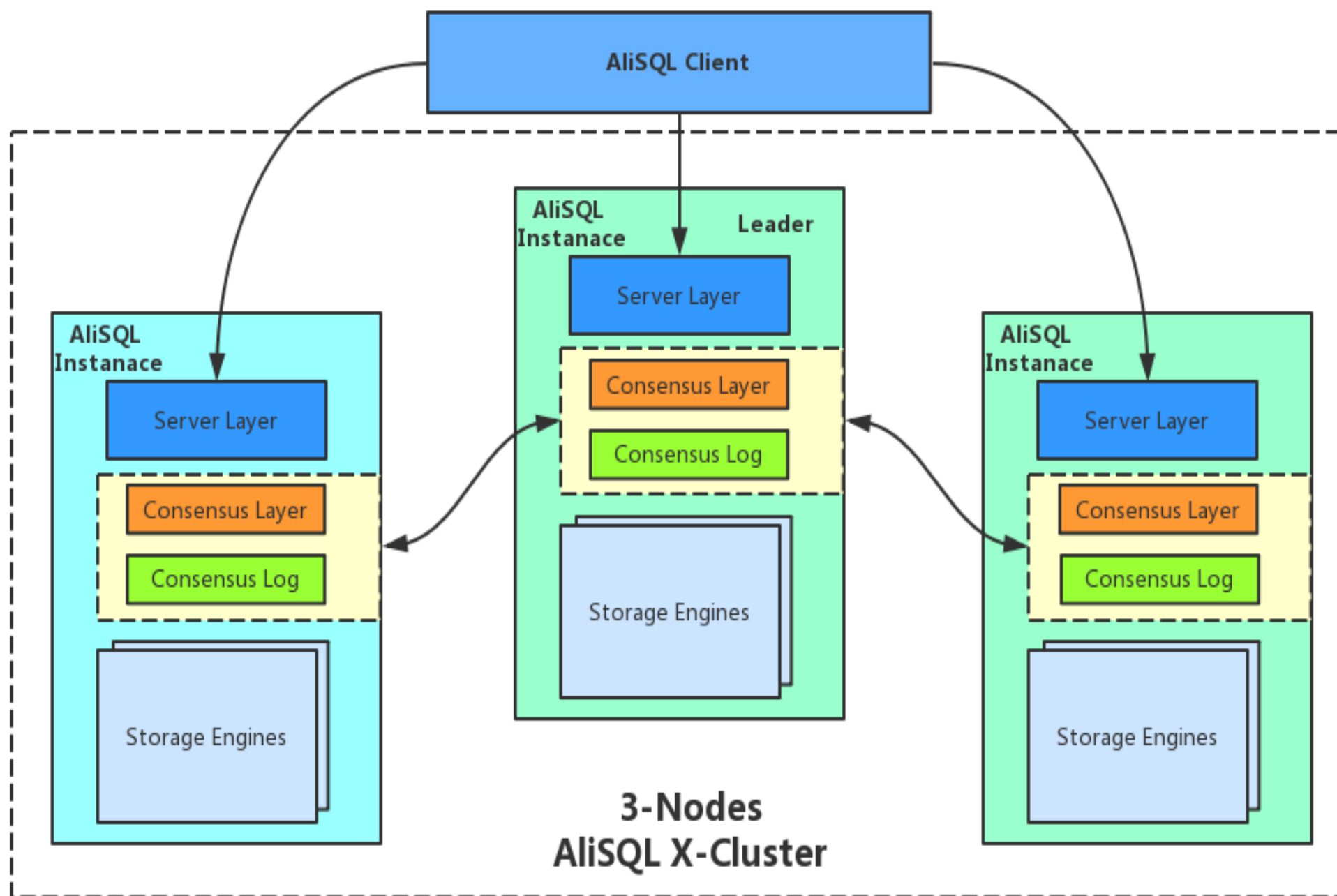
Our Problems

Huge Network Latency!!!
Performance ?



What is X-Cluster

AliSQL Cluster With high availability and strong consistency

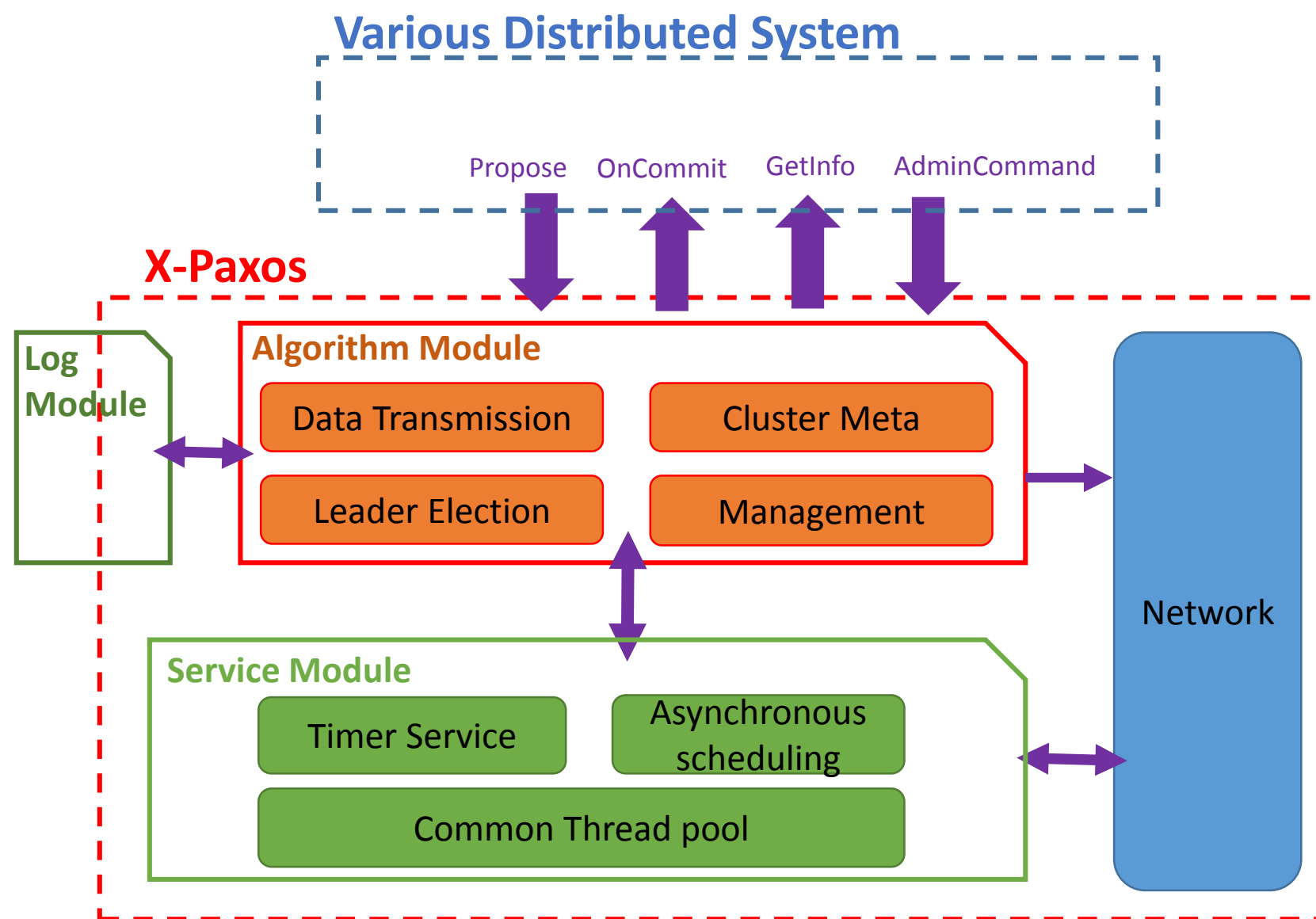


X-Paxos instead of M-S Replication

X-Paxos

Consensus Library With Extreme High Performance And Availability

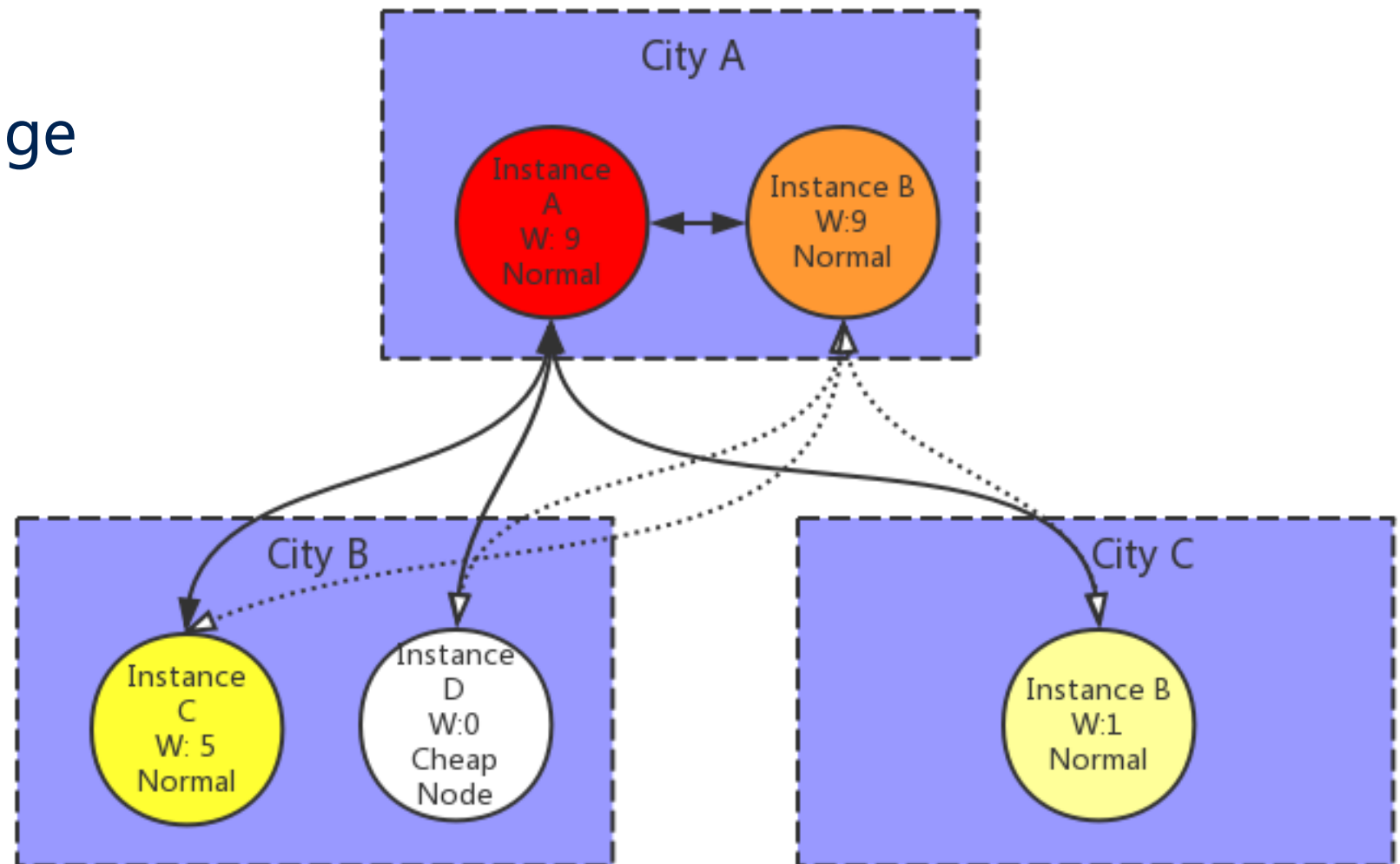
- Multi Paxos Algorithm
- Async Framework
- Pluggable Log



X-Paxos

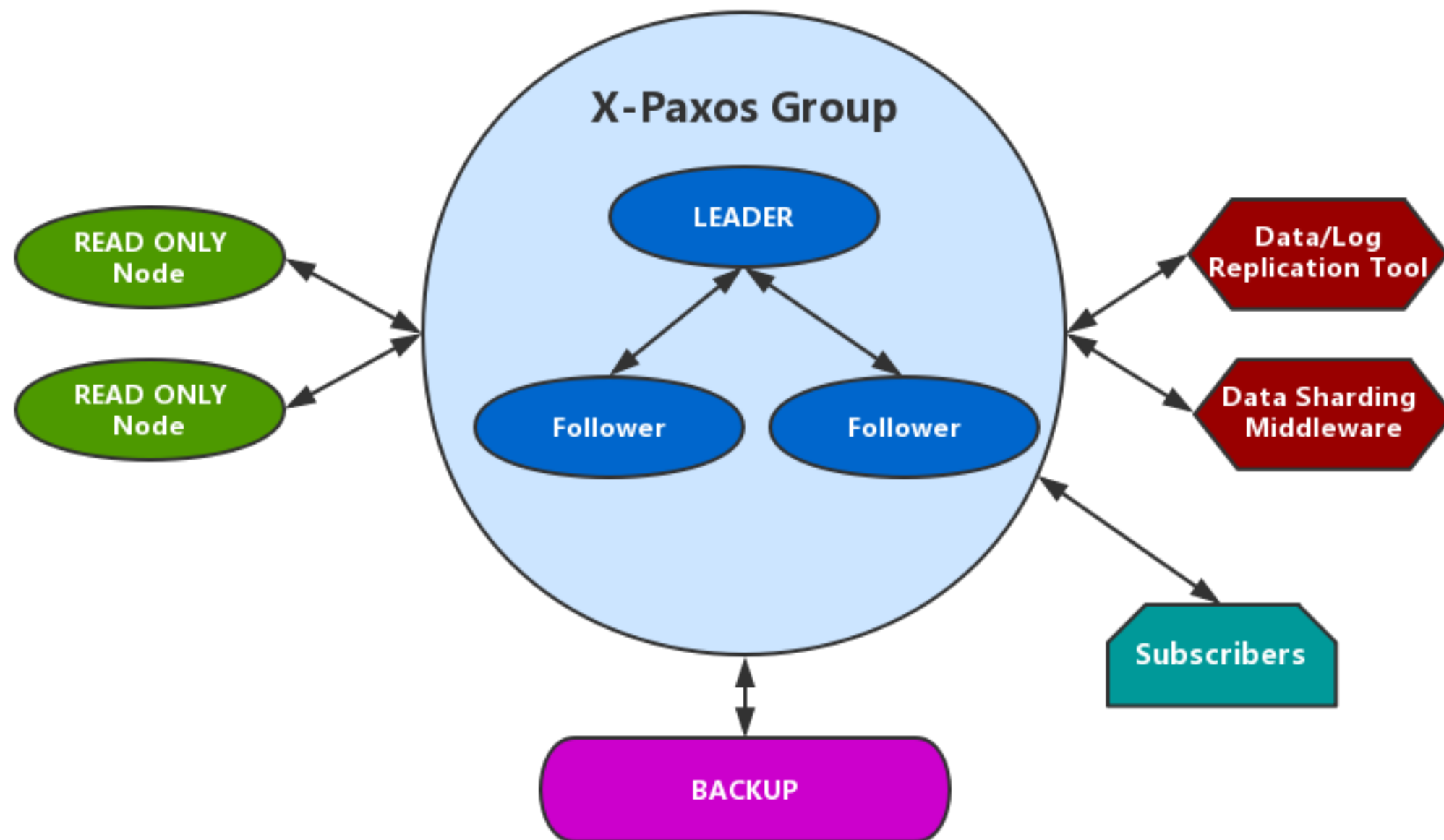
Features Description

- Online Member Change
- Election Weights
- Customized Role



X-Cluster System With X-Paxos

Witness SDK



X-Cluster

- X Cluster uses X-Paxos Pluggable log feature, and design the consensus log with binlog format.
- Consensus Log = Binlog + Relay Log, One instance One Consensus log.
- Transaction is permitted to commit only if majority of instances in the cluster synchronized the log.

Optimization on Performance

Log Replication & Server layer



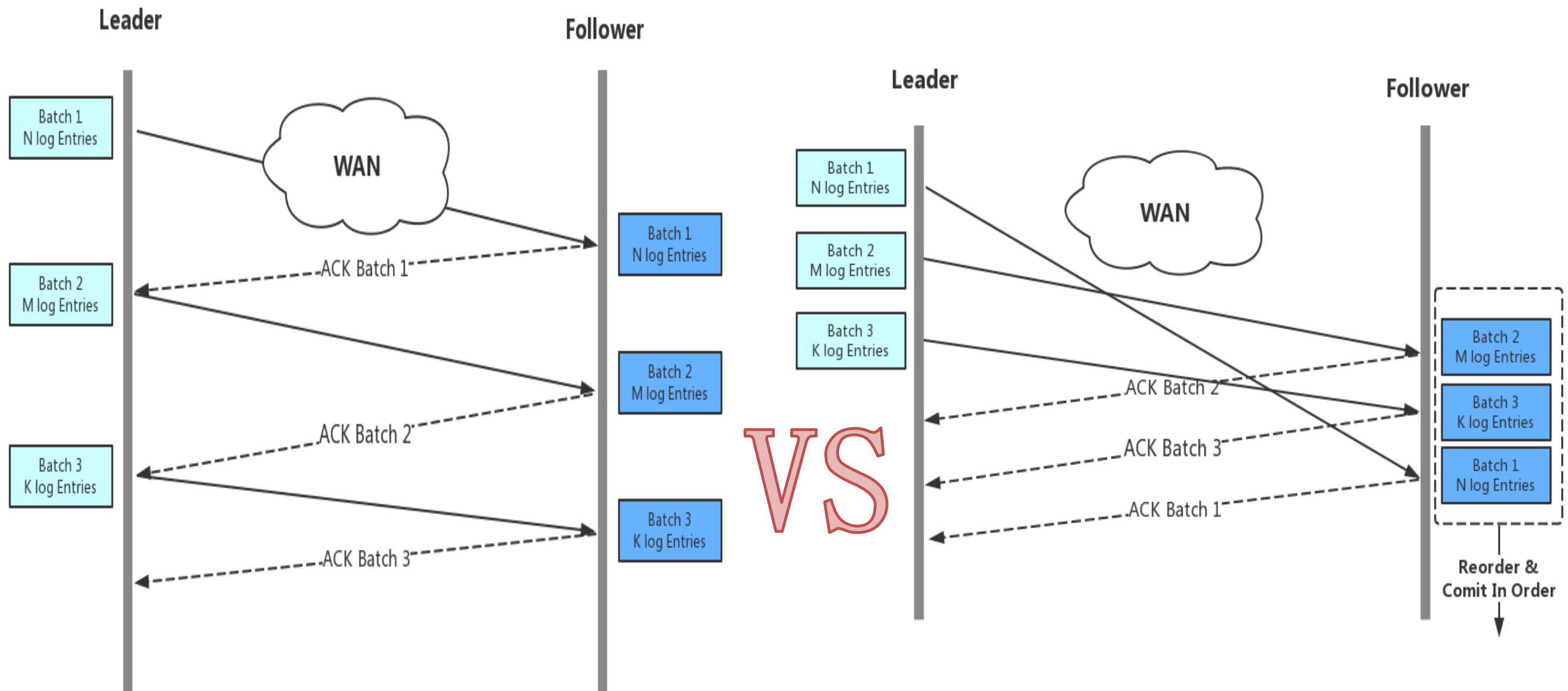
Batching

Pipelining



Optimization on X-Paxos

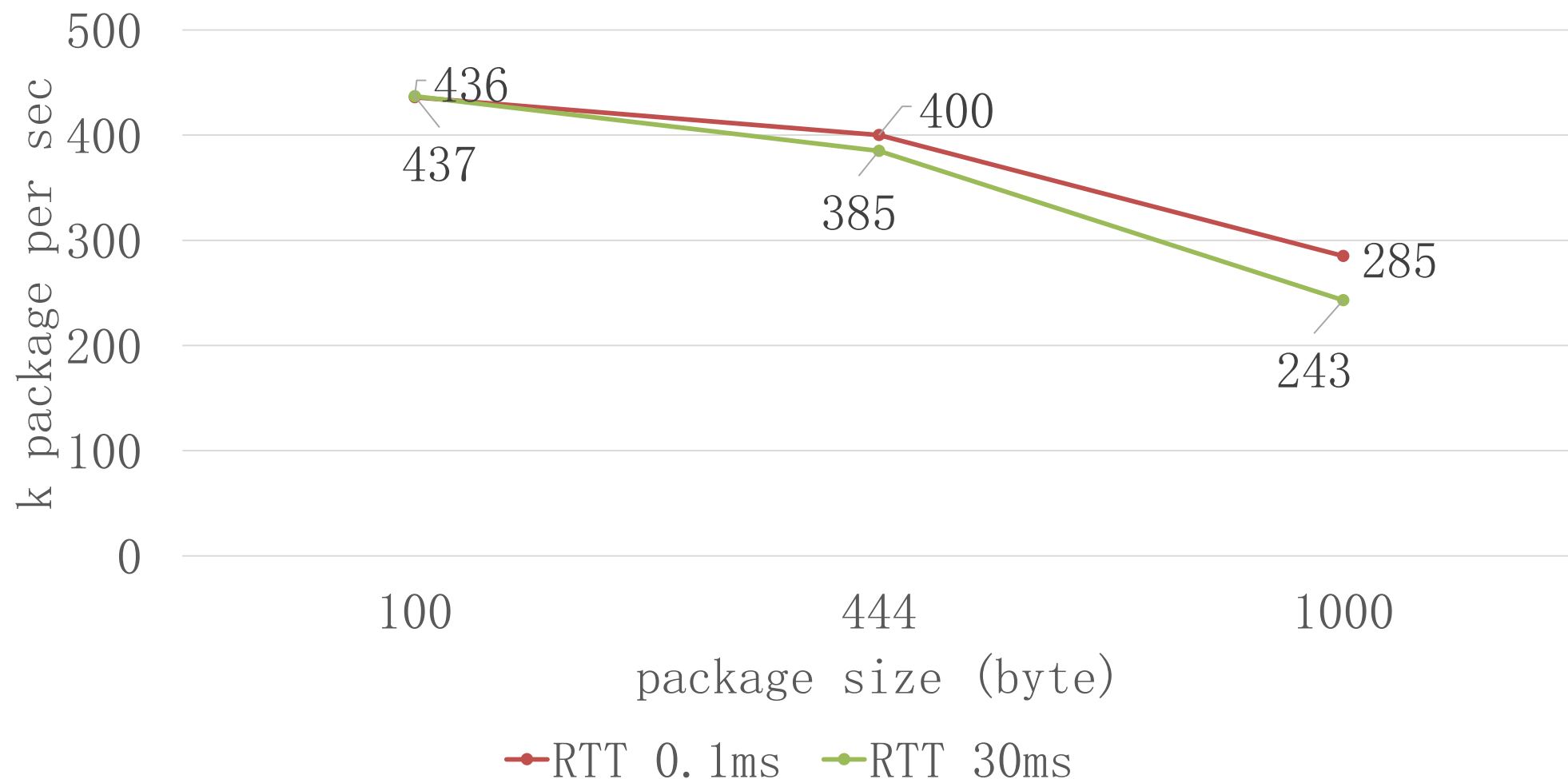
Log Replication batching & Pipelining



Performance of X-Paxos

Low latency & High latency

Replication Benchmark



Throughput of X-Paxos is almost same in low and high latency network

Optimization of X-Cluster

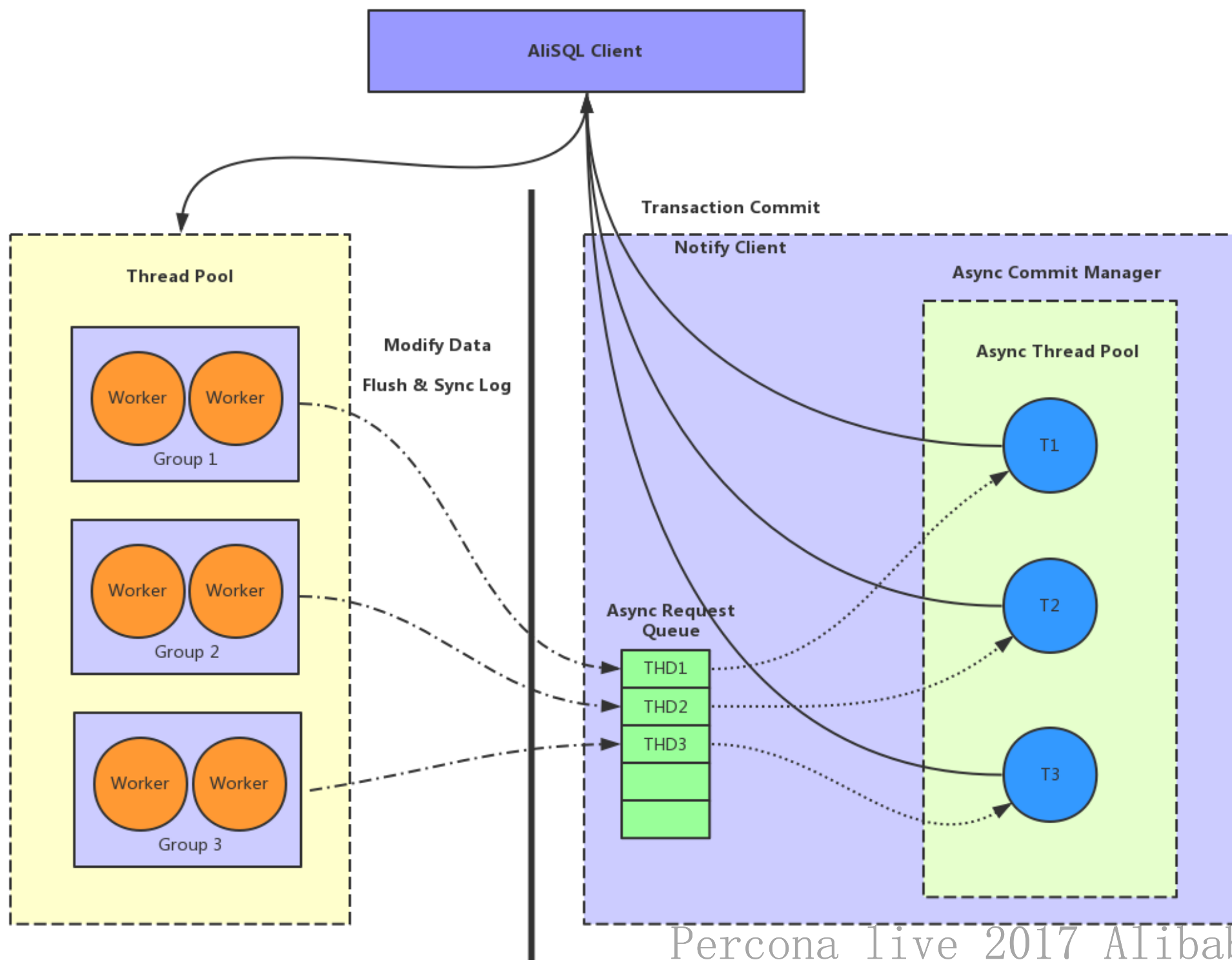
Another problem in High Latency network

Thread pool limit the number of concurrent worker threads to eliminate the data race, but most of the threads are waiting for Committing !!!



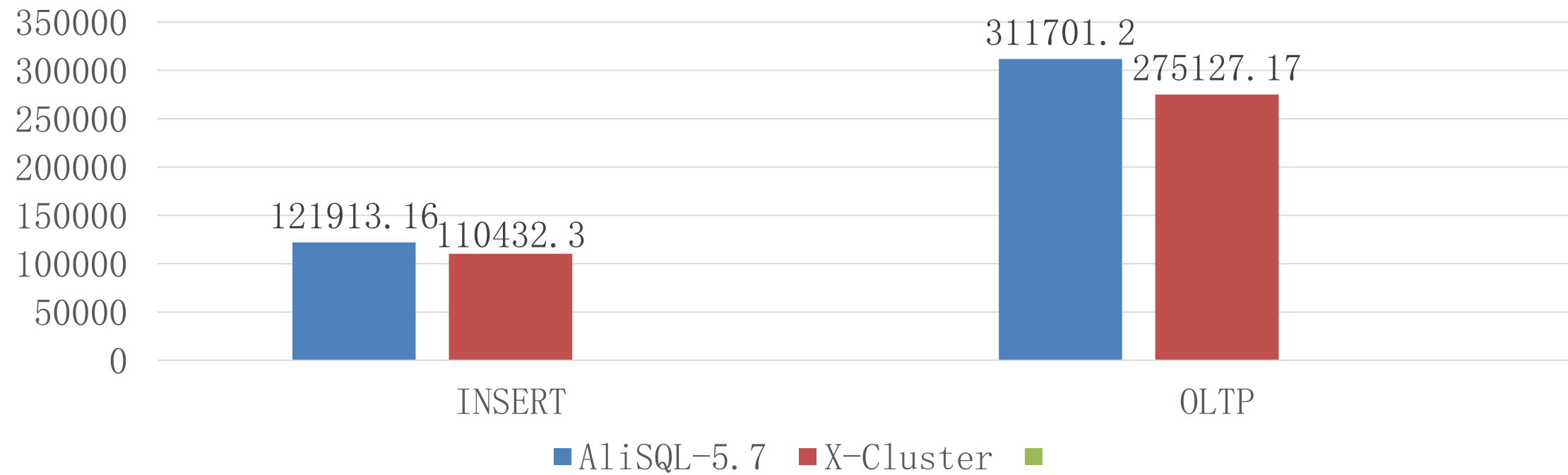
Asynchronously Commit

Commit Thread pool

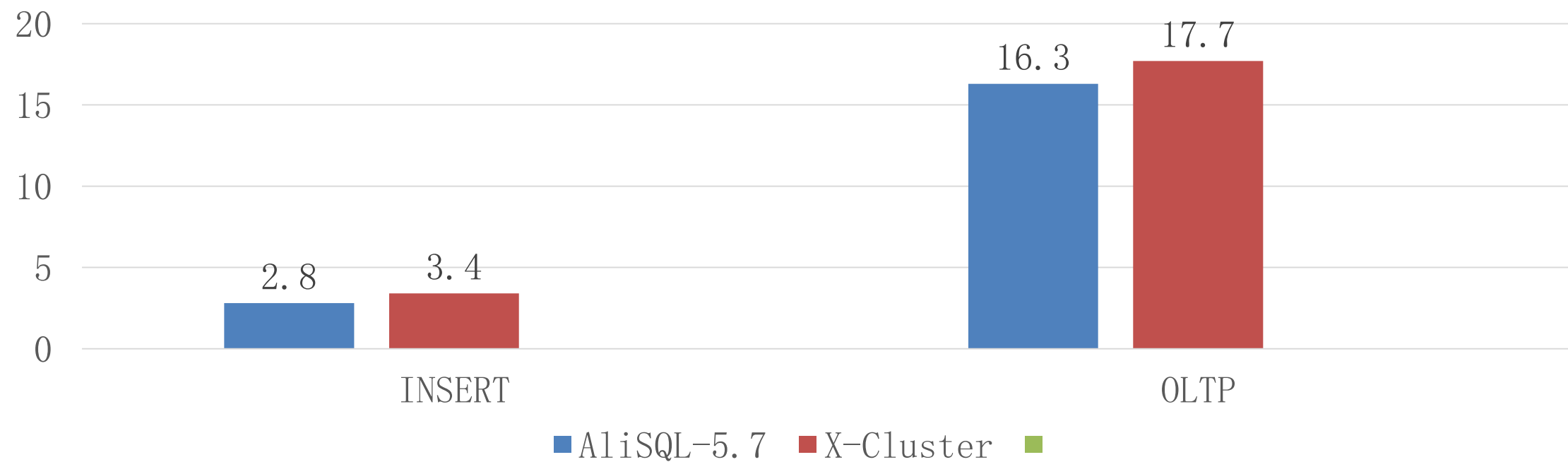


Benchmark

Sysbench Insert & Oltp TPS Performance

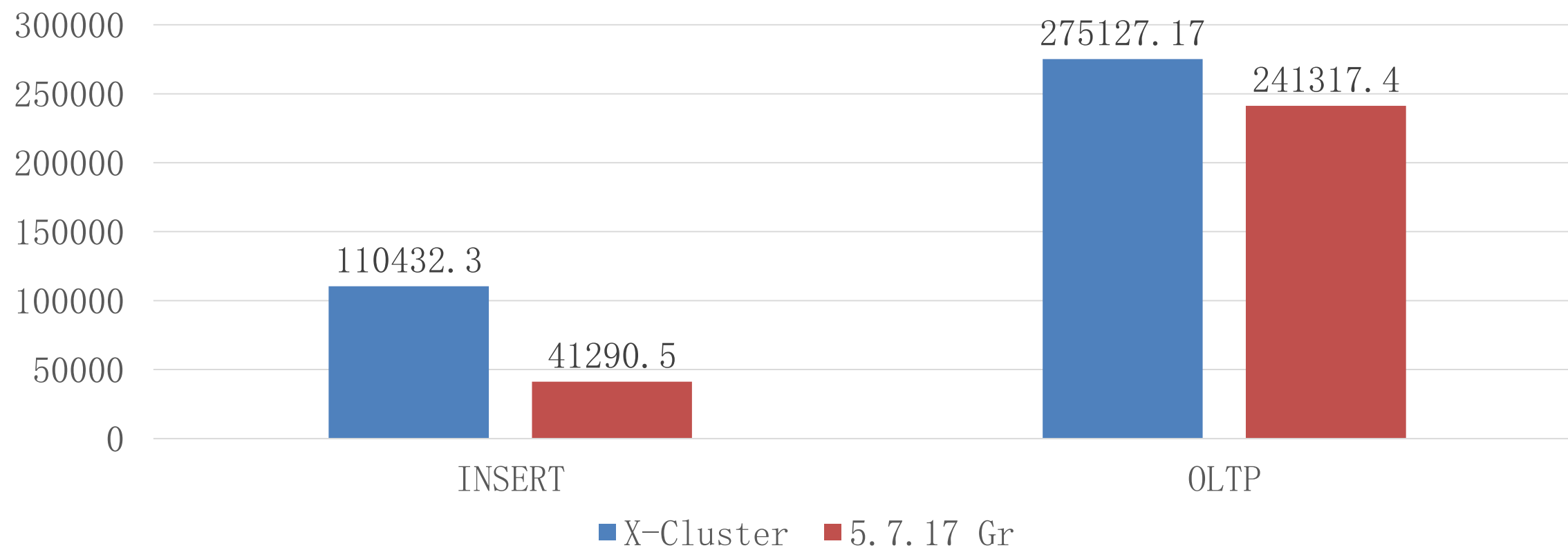


Sysbench Insert & Oltp RT Performance

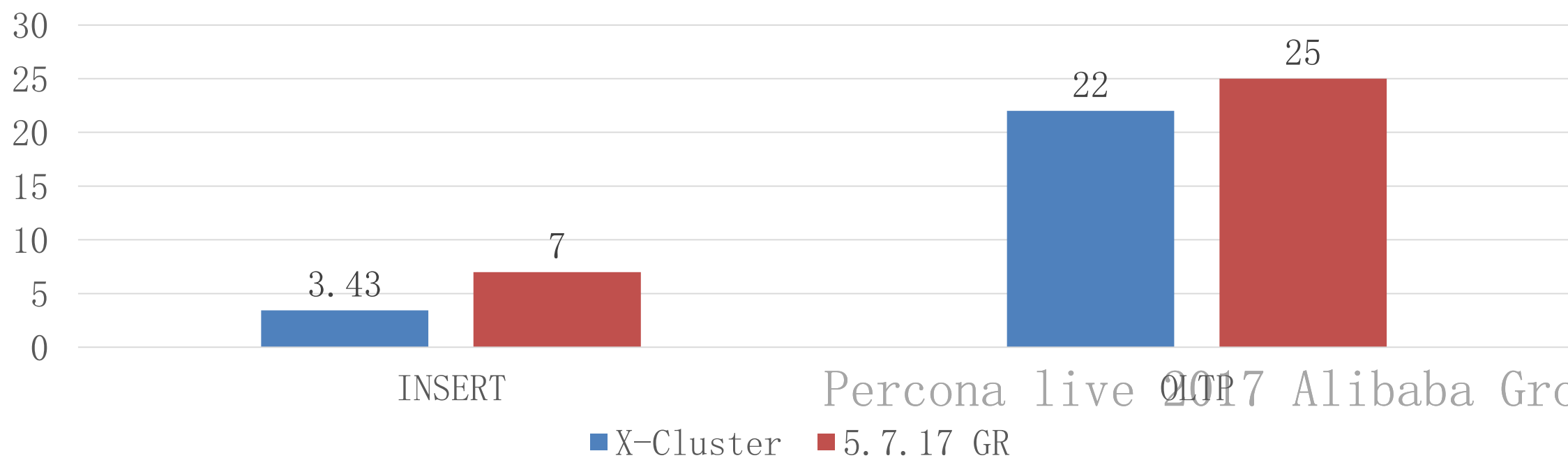


X-Cluster vs GR

Sysbench QPS in low latency(0.1ms) network

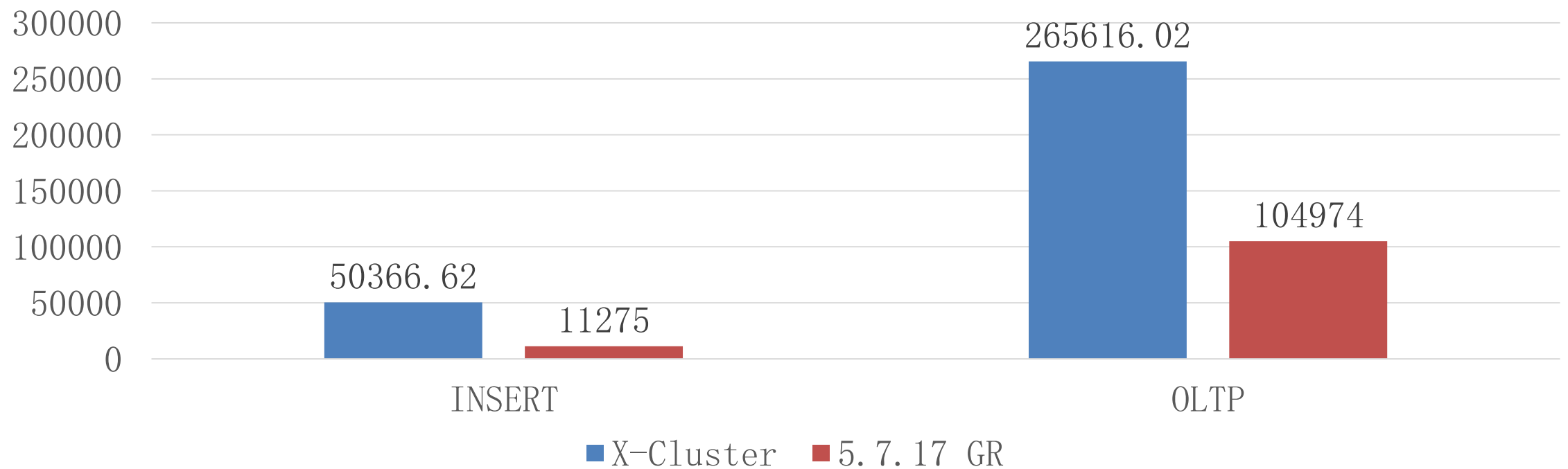


Sysbench RT in low latency(0.1ms) network

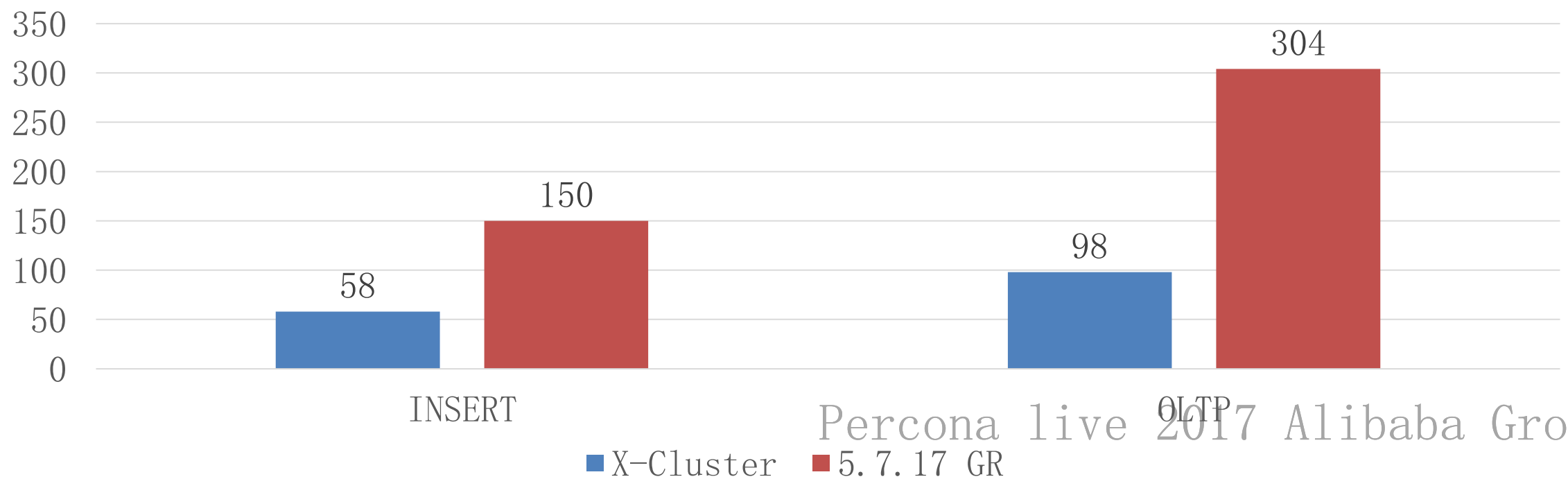


X-Cluster vs GR

Sysbench QPS in high latency(30ms) network



Sysbench RT in high latency(30ms) network



Summary

X-KV and X-Cluster is Our new solution base AliSQL

■ X-KV

- More data types
- New protocol
- Better performance
- Easy maintenance

■ X-Cluster

- Strong consistency
- Geo distributed
- High performance

Wei Hu

Database Kernel Developer @ Alibaba Group

Contact info:

Email: droopy.hw#alibaba-inc.com

Facebook: Droopy Hu

Q&A

Thank you

Alibaba Database Team