

# 如何实现一个工业级的Paxos协议库

章颖强（江疑）





## 关于我

- 章颖强（花名江疑），阿里巴巴数据库事业部高级专家，拥有10+年的分布式系统、数据库内核研发经验。曾任AliSQL项目研发负责人，多次担任阿里双十一数据库内核负责人，阿里分布式强一致协议X-Paxos作者。
- 目前为阿里新一代数据库X-DB项目分布式与基础方向负责人，专注于分布式协议、基础技术等方向。





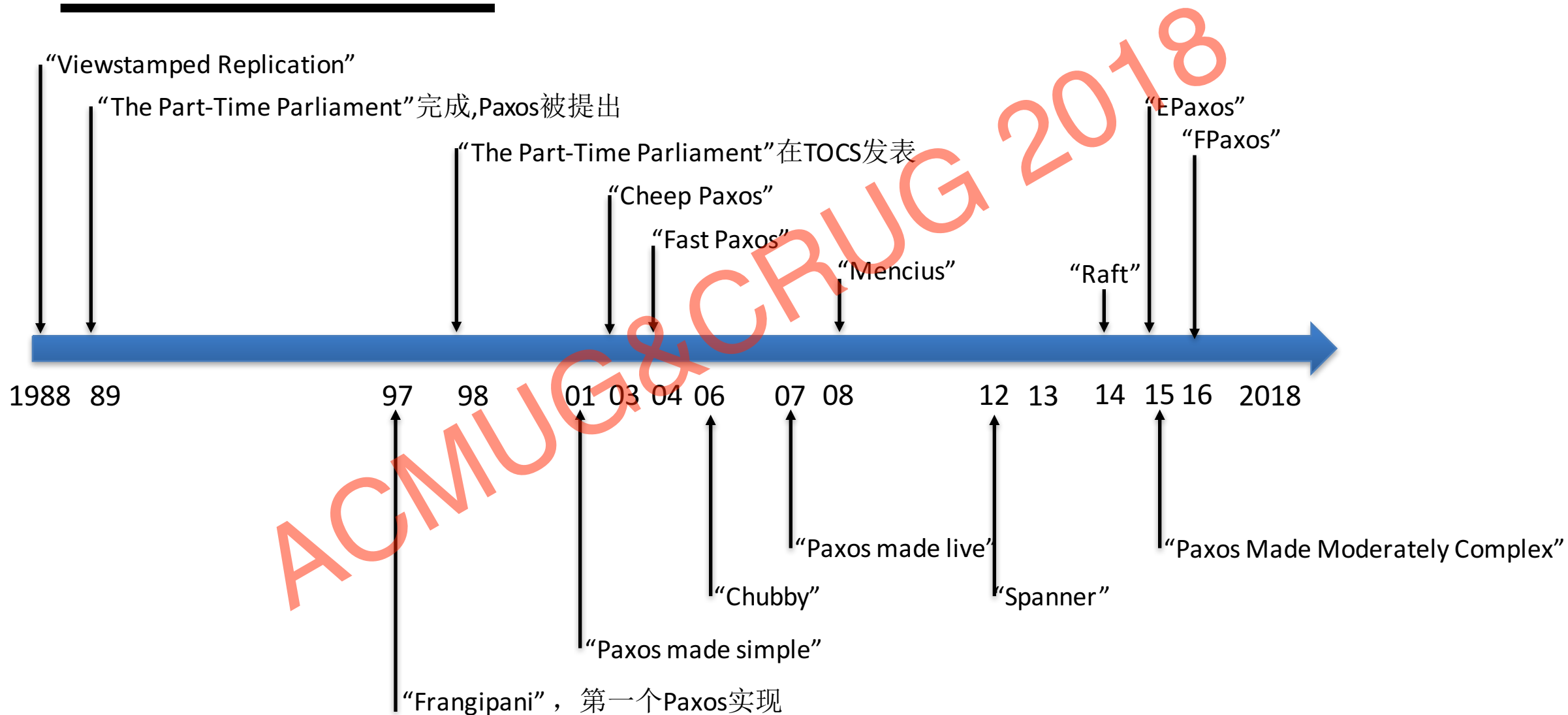
什么是Paxos

为什么说X-Paxos是工业级的

X-Paxos的应用



# Paxos的历史





# 什么是Paxos

---

## □ 理论视角

- ✓ Paxos is a family of protocols for solving **consensus** in a network of unreliable processors.
- ✓ **Consensus** is the process of agreeing on one result among a group of participants.

## □ 实用视角

- ✓ Consensus protocols(Paxos) are the basis for the **state machine replication(SMR)** approach to distributed computing
- ✓ **State machine replication** is a technique for converting an algorithm into a fault-tolerant, distributed implementation.



# 当前Paxos的前沿研究

## □ 理论突破

- ✓ Flexible Paxos: Quorum Intersection Revisited - OPODIS 2016
- ✓ XFT- Practical Fault Tolerance beyond Crashes (XPaxos) – OSDI 2016
- ✓ WPaxos: Ruling the Archipelago with Fast Consensus – CSDC 2017

## □ 软硬件结合

- ✓ Just Say NO to Paxos Overhead: Replacing Consensus with Network Ordering – OSDI 2016
- ✓ APUS- Fast and Scalable PAXOS on RDMA – SoCC 2017

## □ 应用

- ✓ PaxosStore- High-availability Storage Made Practical in WeChat - VLDB 2017
- ✓ Protocol-Aware Recovery for Consensus-Based Storage – FAST 2018



什么是Paxos

什么是工业级的Paxos

X-Paxos的应用



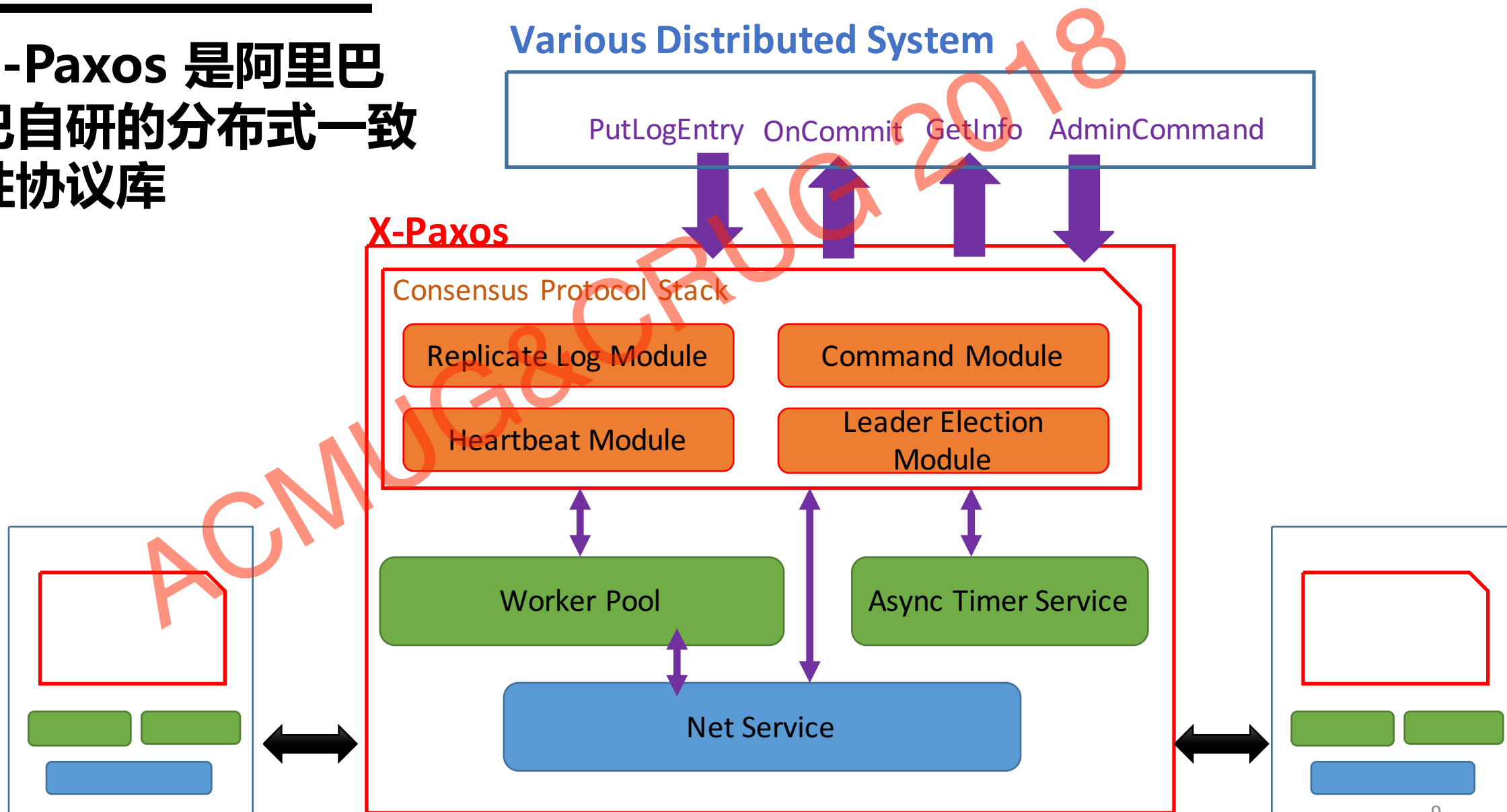
# 工业级Paxos的要求





# 什么是X-Paxos

- X-Paxos 是阿里巴巴自研的分布式一致性协议库





# X-Paxos的核心技术——全异步框架

## □ 自研基于Future/Promise思想的全异步化调度框架

- ✓ 并行流水线
- ✓ 内存拷贝降低
- ✓ 用户态调度

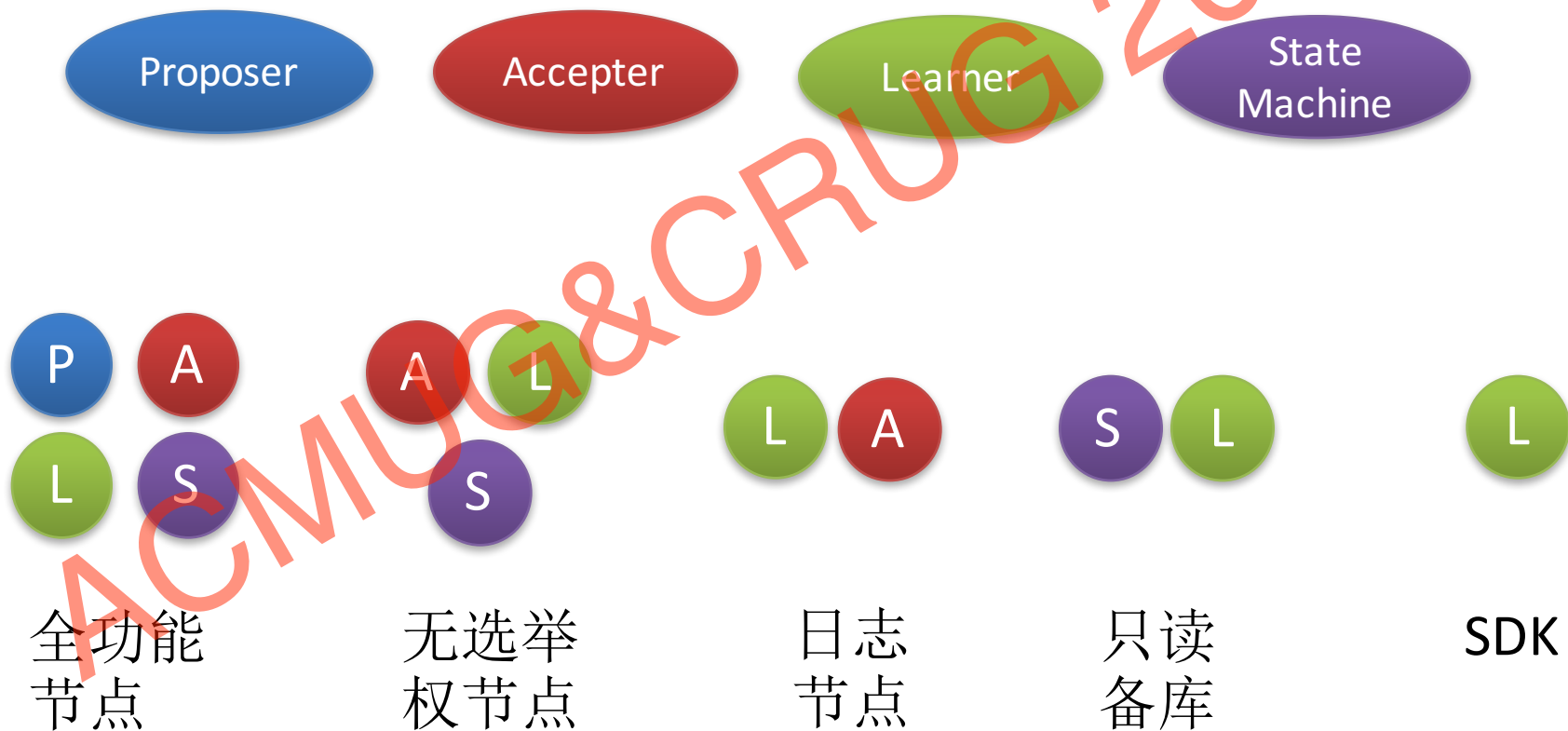
## □ 基于C++14新特性

- ✓ Move语义
- ✓ 泛型设计



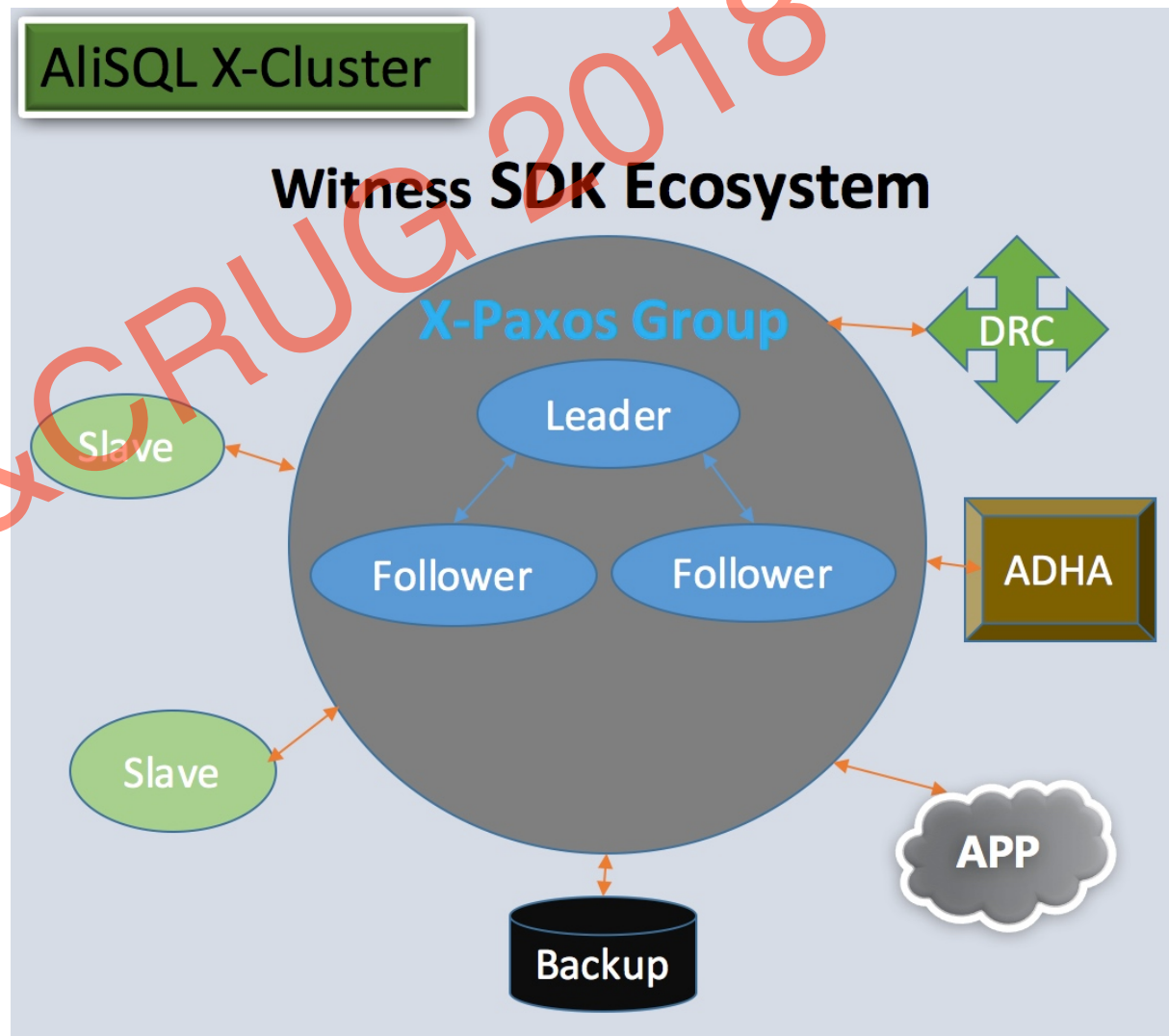


# X-Paxos的核心技术——动态定制角色



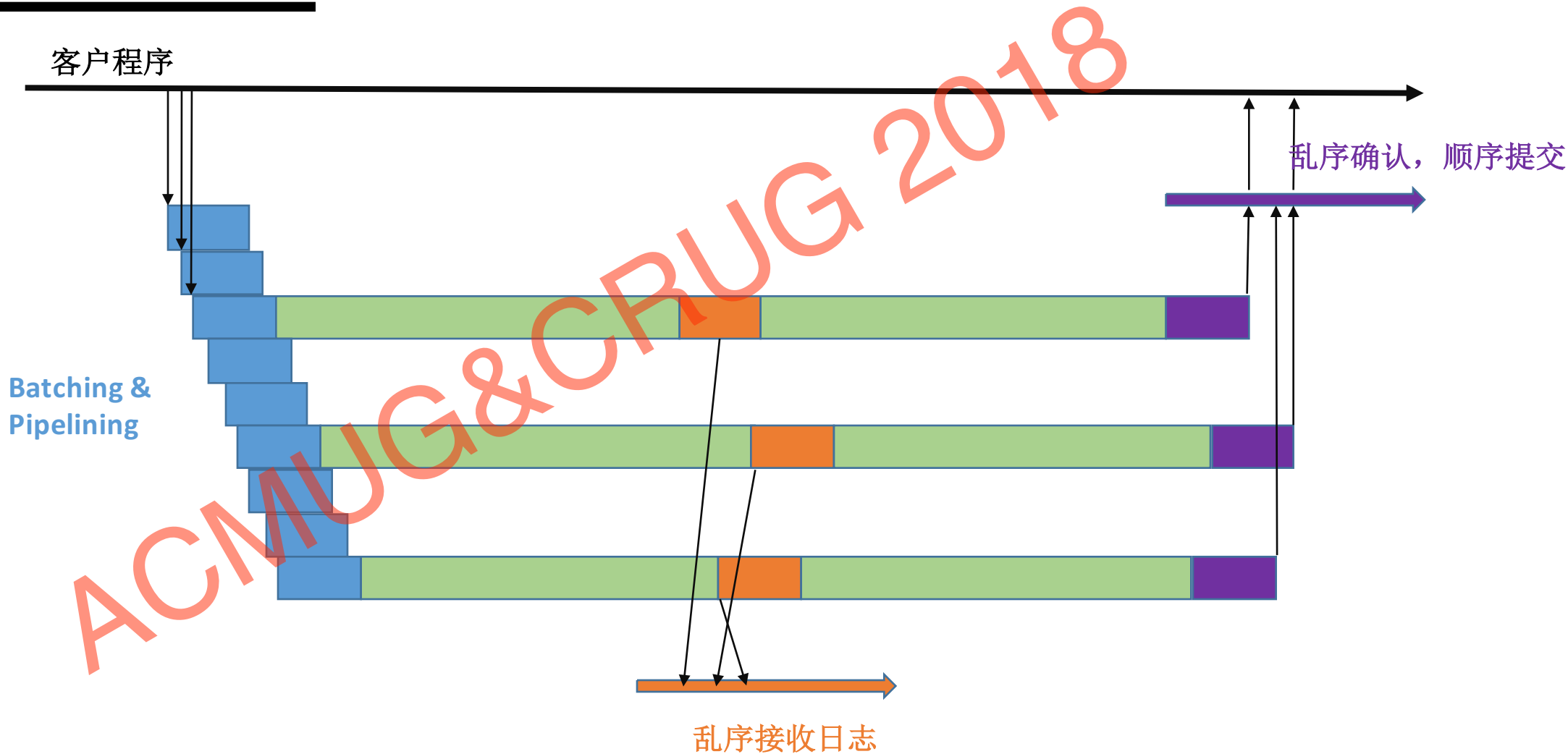
# X-Paxos的核心技术——SDK

- 基于X-Paxos SDK  
的统一生态系统



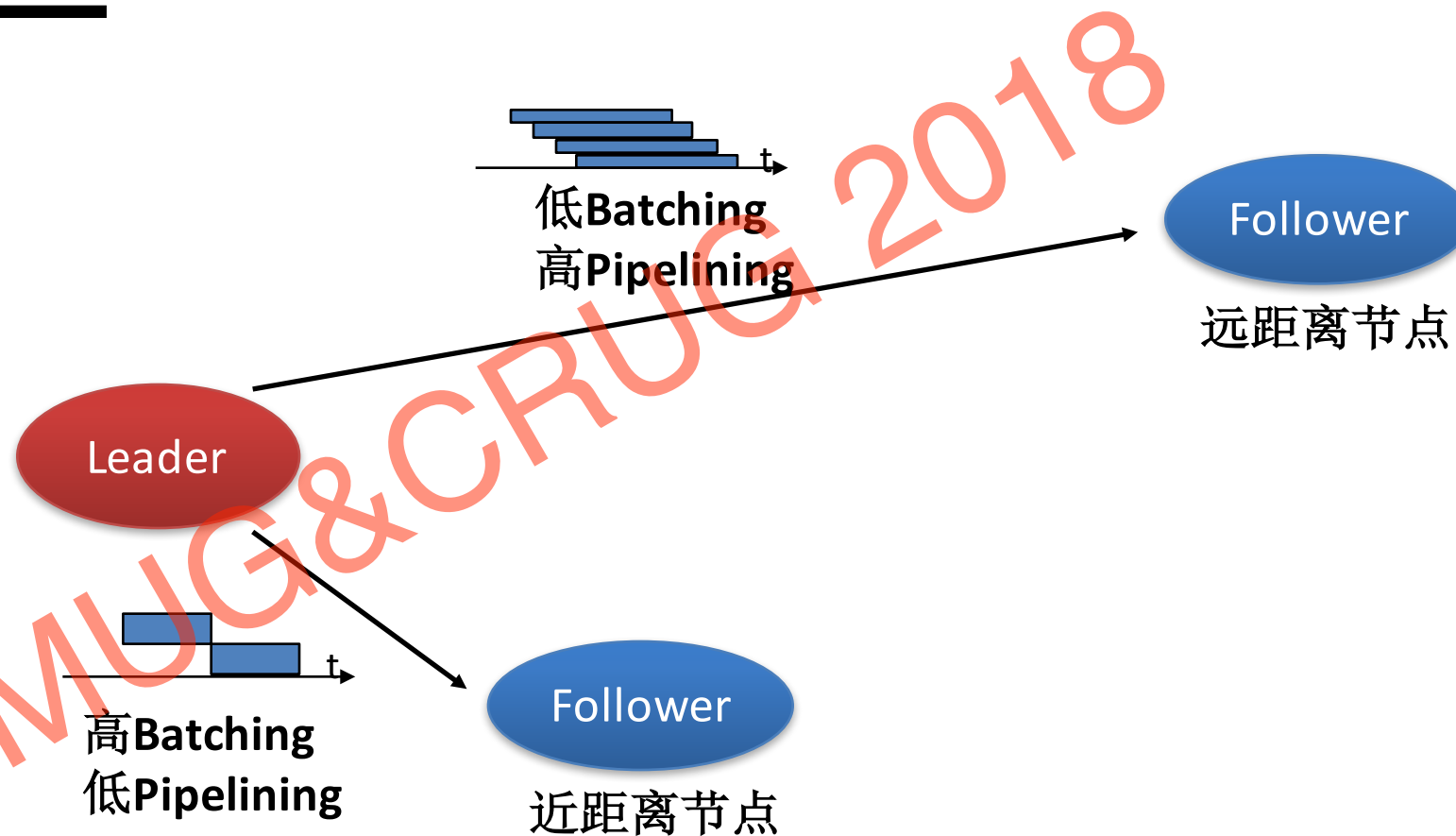


# X-Paxos的核心技术——自适应传输优化

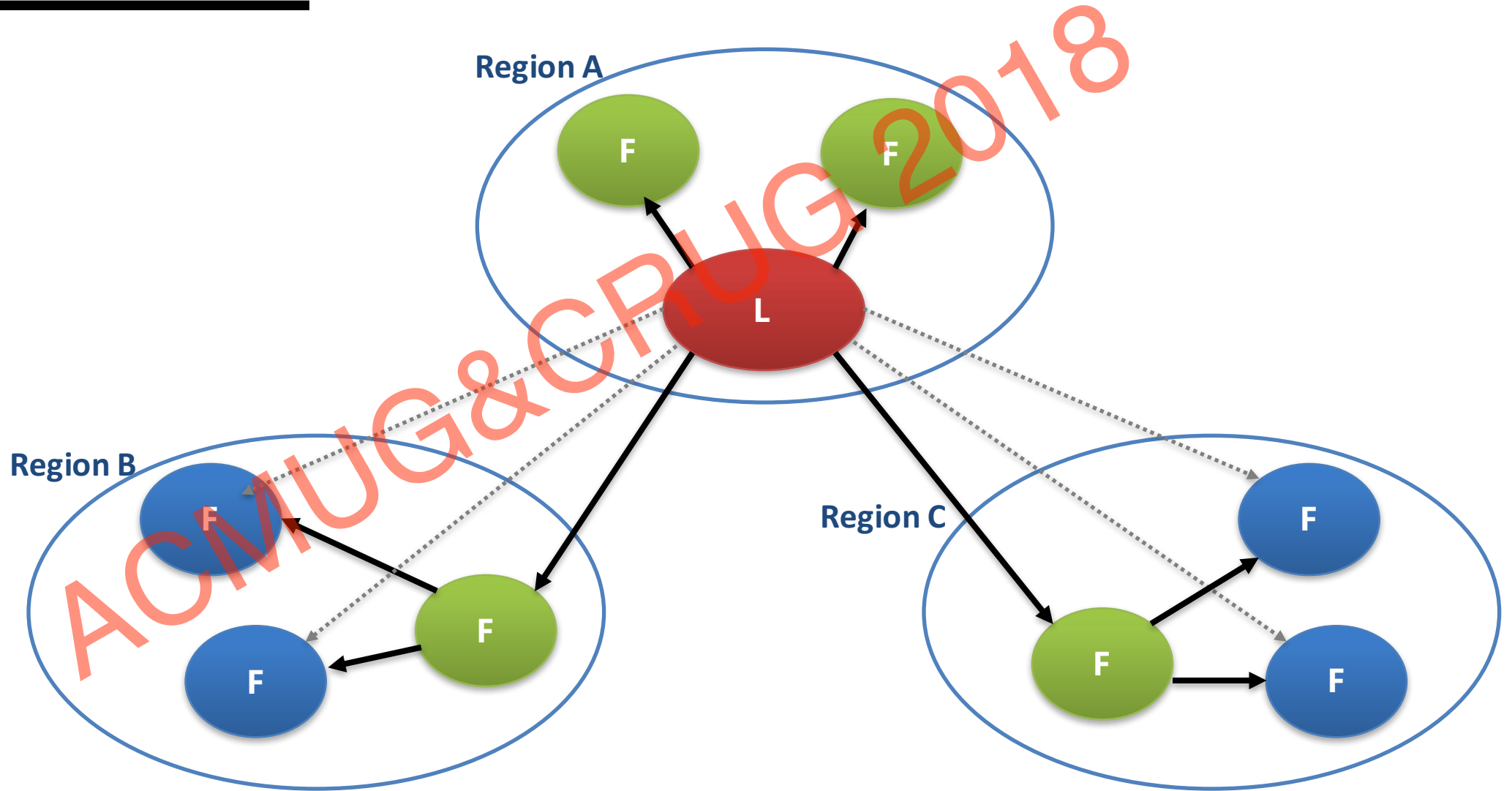




# X-Paxos的核心技术——自适应传输优化



# X-Paxos的核心技术——LACD





# X-Paxos的核心技术——可插拔日志

## □ 完善的日志接口设计

✓ 快速集成

## □ 统一的日志实现

✓ 高性能

### X-Paxos

#### Consensus Protocol Stack

Transmission Module

Election module

Meta Module

Controller

PaxosLog Interface

继承

### X-DB

WAL/BINLOG

Server Layer

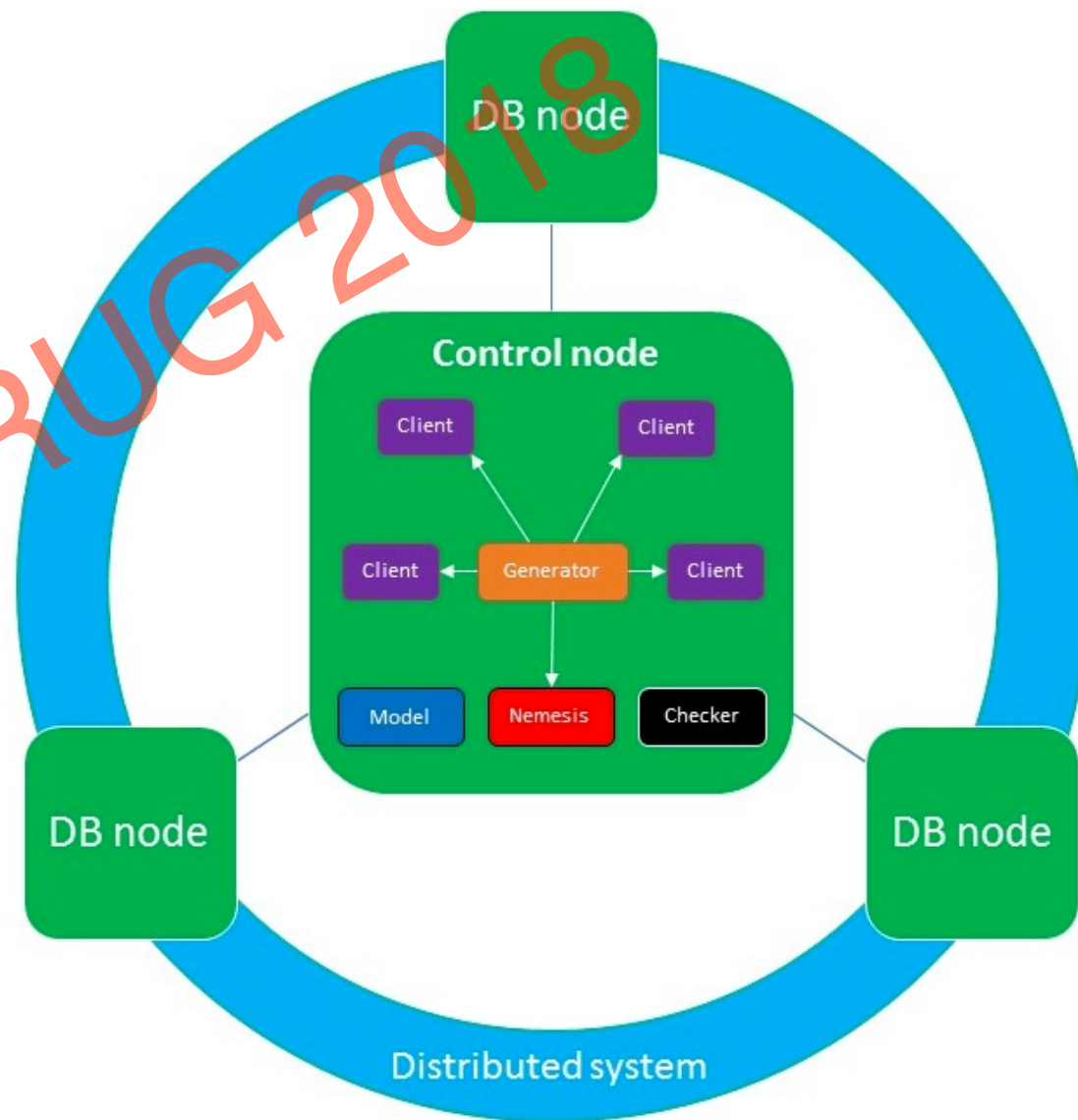
Storage Engine



# X-Paxos的核心技术——完备的测试



# X-Paxos的核心技术——Jepsen





# 工业级Paxos vs X-Paxos

性能

全异步框架

无锁设计

自适应传输优化

功能

动态定制角色

权重选主

扩展性

自适应传输优化

SDK

测试验证

Jepsen

TLA+

线上验证

X-DB 万级部署规模

多个线上业务验证



什么是Paxos

为什么说X-Paxos是工业级的

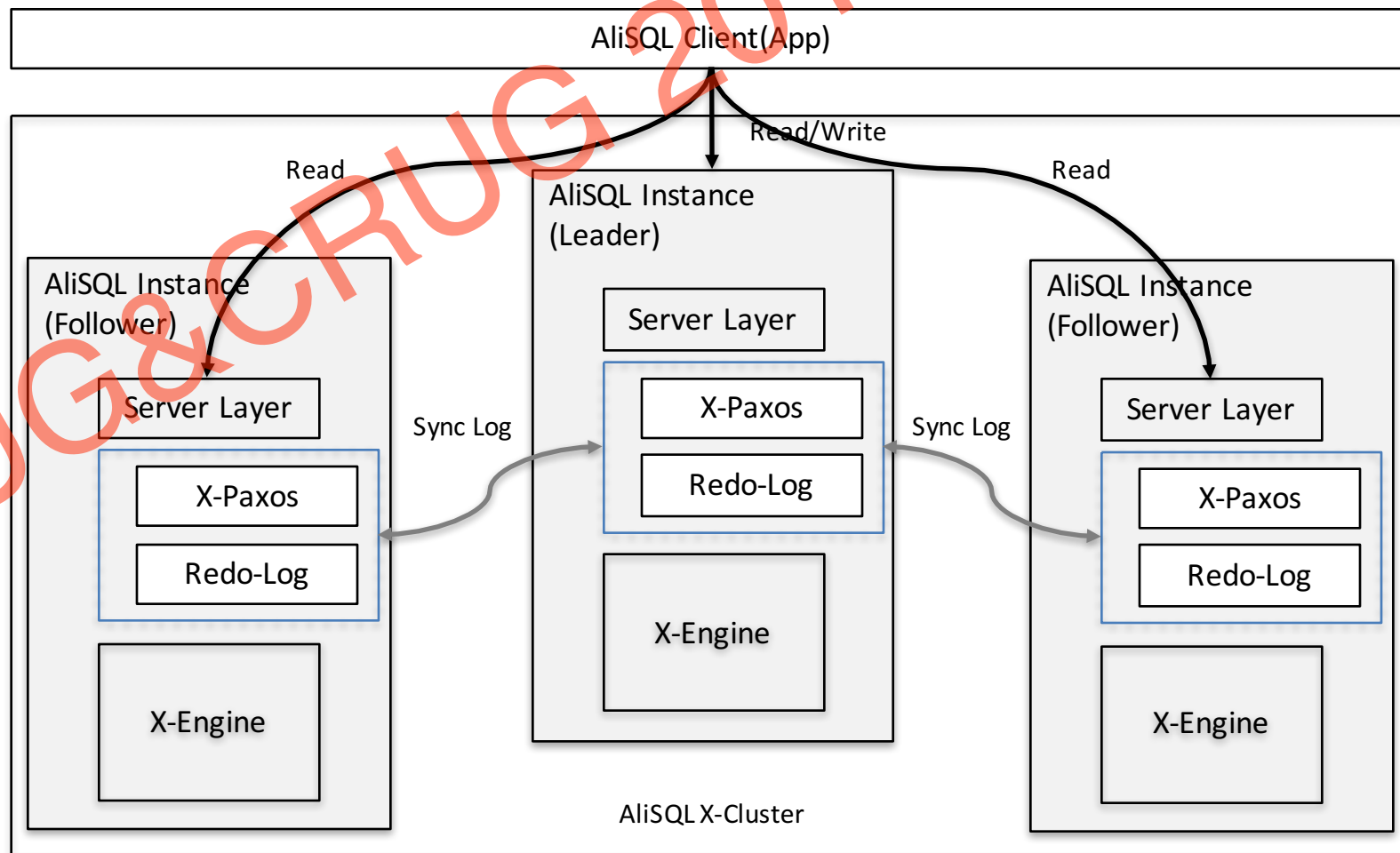
X-Paxos的应用



# X-Paxos在阿里的应用——X-DB 1.0

## □ X-DB 1.0 ( 原X-Cluster )

- ✓ 能力：金融级可全球化部署数据库

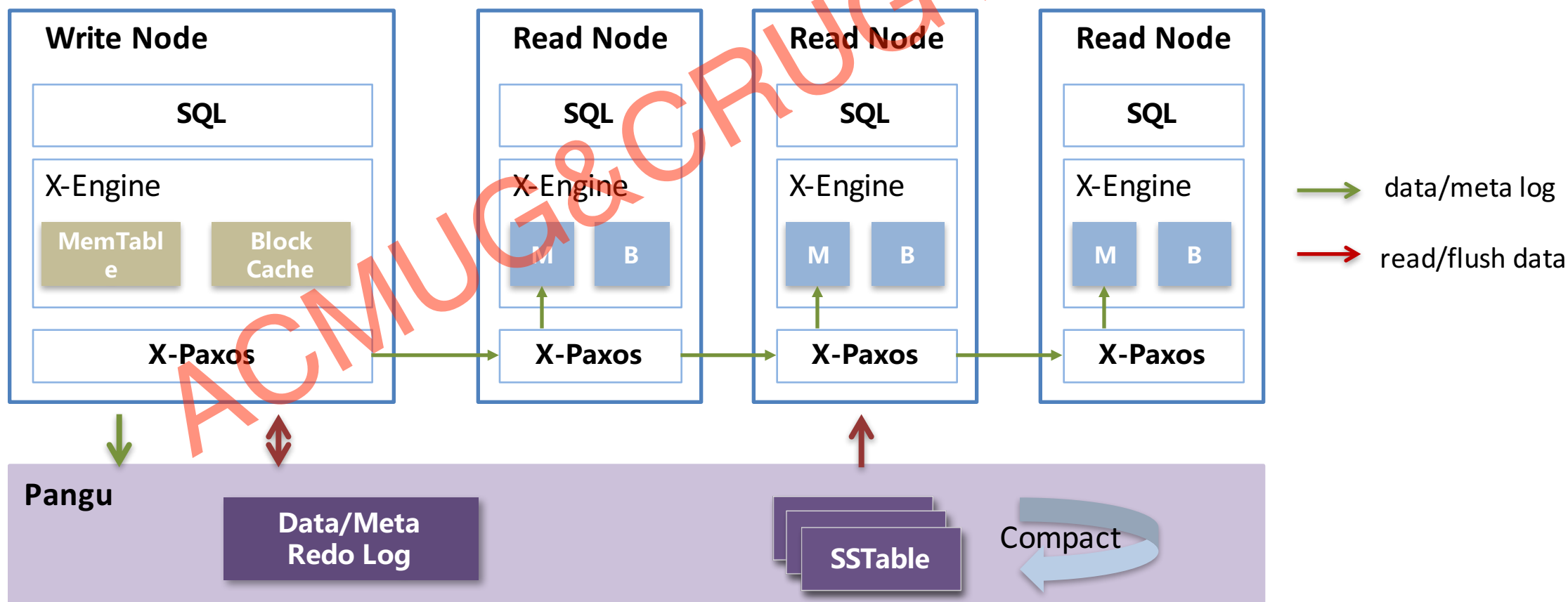




# X-Paxos在阿里的应用——X-DB 2.0

## □ X-DB 2.0 (一写多读)

- ✓ 能力：基于盘古共享存储系统，提供用户可水平扩展的存储能力以及只读节点的扩展能力
- ✓ 效率：基于软硬件的Co-Design，提供超越竞品10倍以上的硬件使用效率

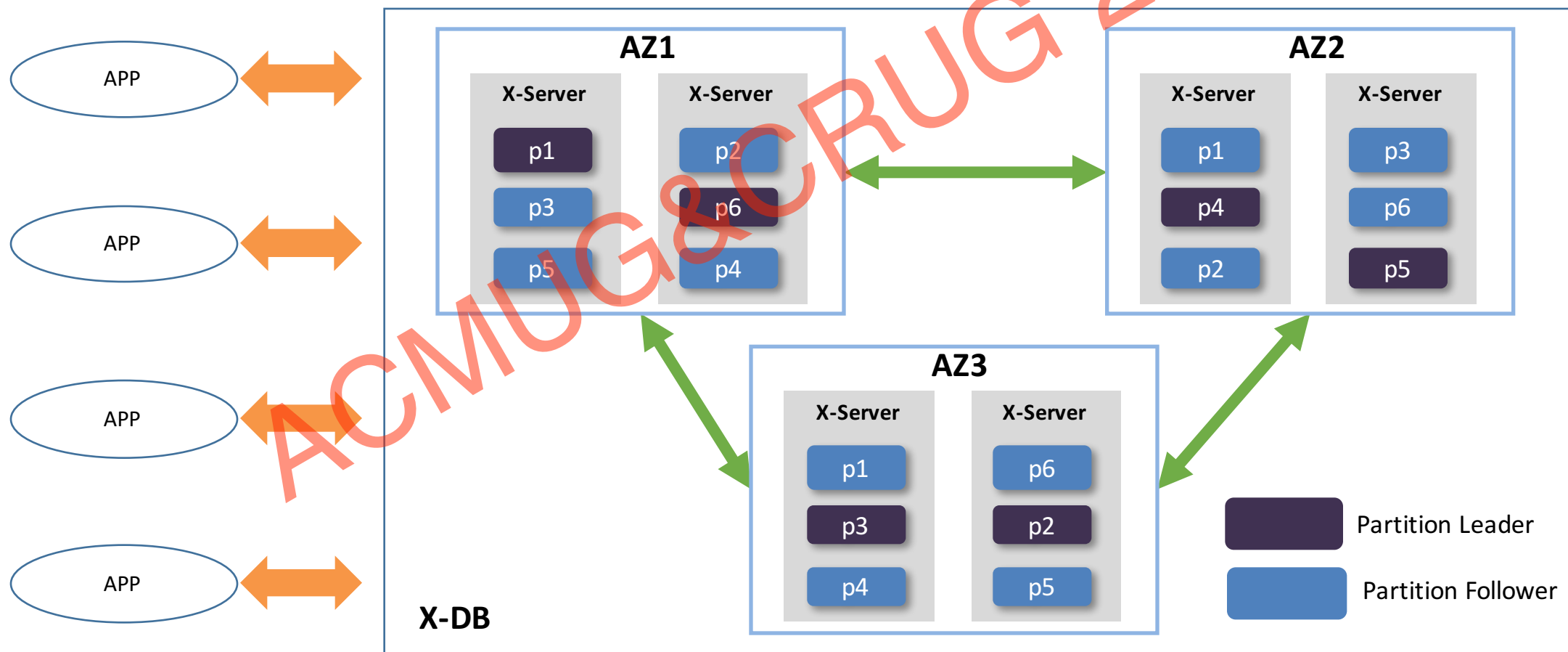




# X-Paxos在阿里的应用——X-DB 3.0

## □ X-DB3.0 ( 多点可读可写 )

- ✓ 能力上：计算和存储的无限扩展能力，满足小客户到大客户对数据库的所有需求
- ✓ 效率上：全面兼容MySQL生态，用户无缝迁移。软硬件结合的Co-Design，打造世界范围内超越竞品的极致竞争力





谢谢大家

