



# TDSQL在金融 场景中的实践



腾讯云

harlylei(雷海林) 专家工程师

# 目录

1. 简介
2. 核心特性
3. 分布式实践
4. 部署实践

# 1 TDSQL简介

定位、合作伙伴

TDSQL (Tencent Distributed MySQL) 是腾讯  
针对**金融场景**推出的强一致性，分布式数据库  
集群解决方案。

# 介绍

**Tencent Distributed MySQL**

专有云命名TDSQL

**关系型**

NoShard

完全兼容MySQL和其分支版本

从GB到TB/EB

**分布式**

Shard

解决超大并发、超大数据量的场景

实时数据挖掘

ETL工具

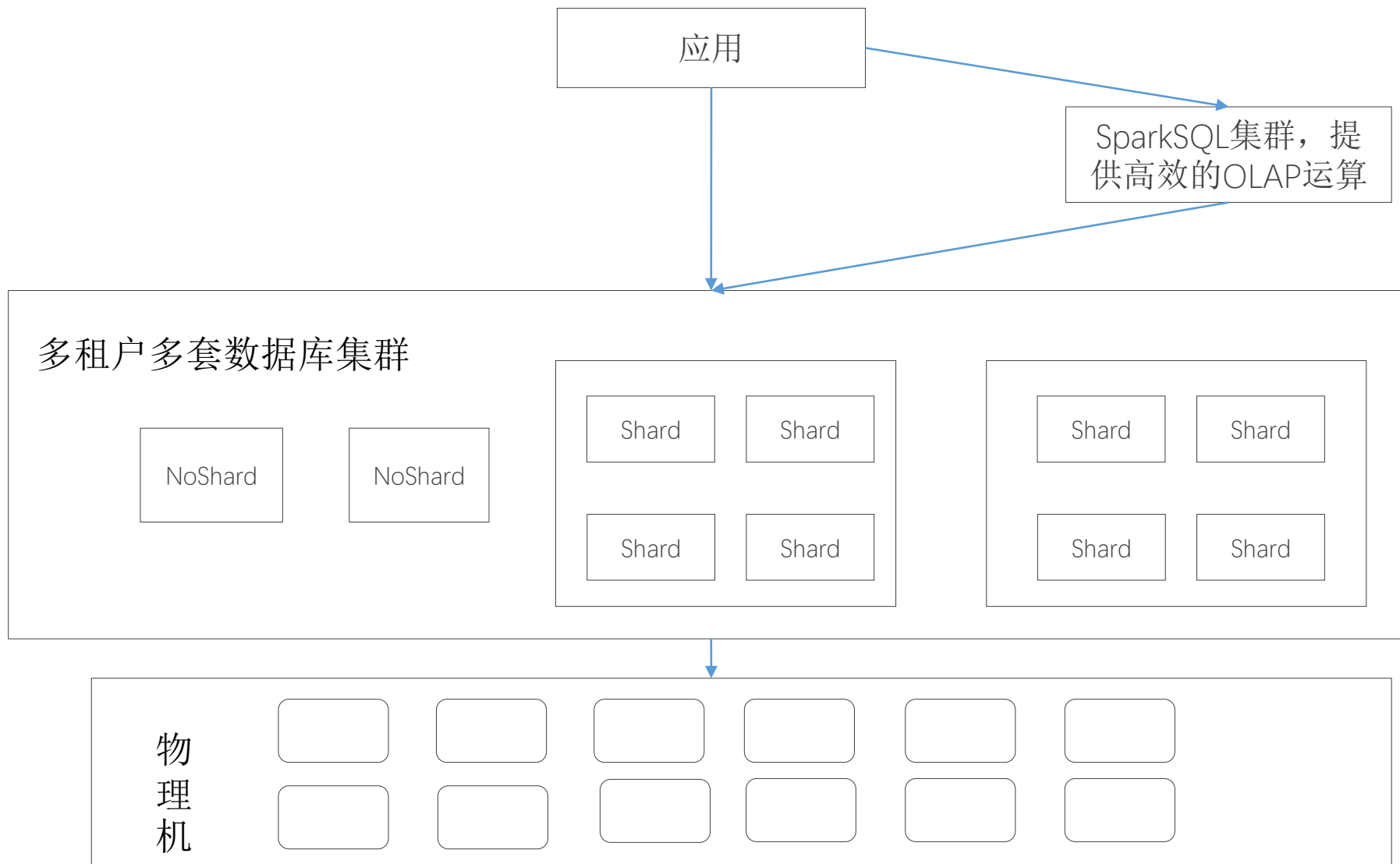
**分析型**

Spark

对接数据分析，大数据平台

公有云命名为 CDB（关系型数据库）DCDB（分布式数据库）

# 灵活的多租户集群



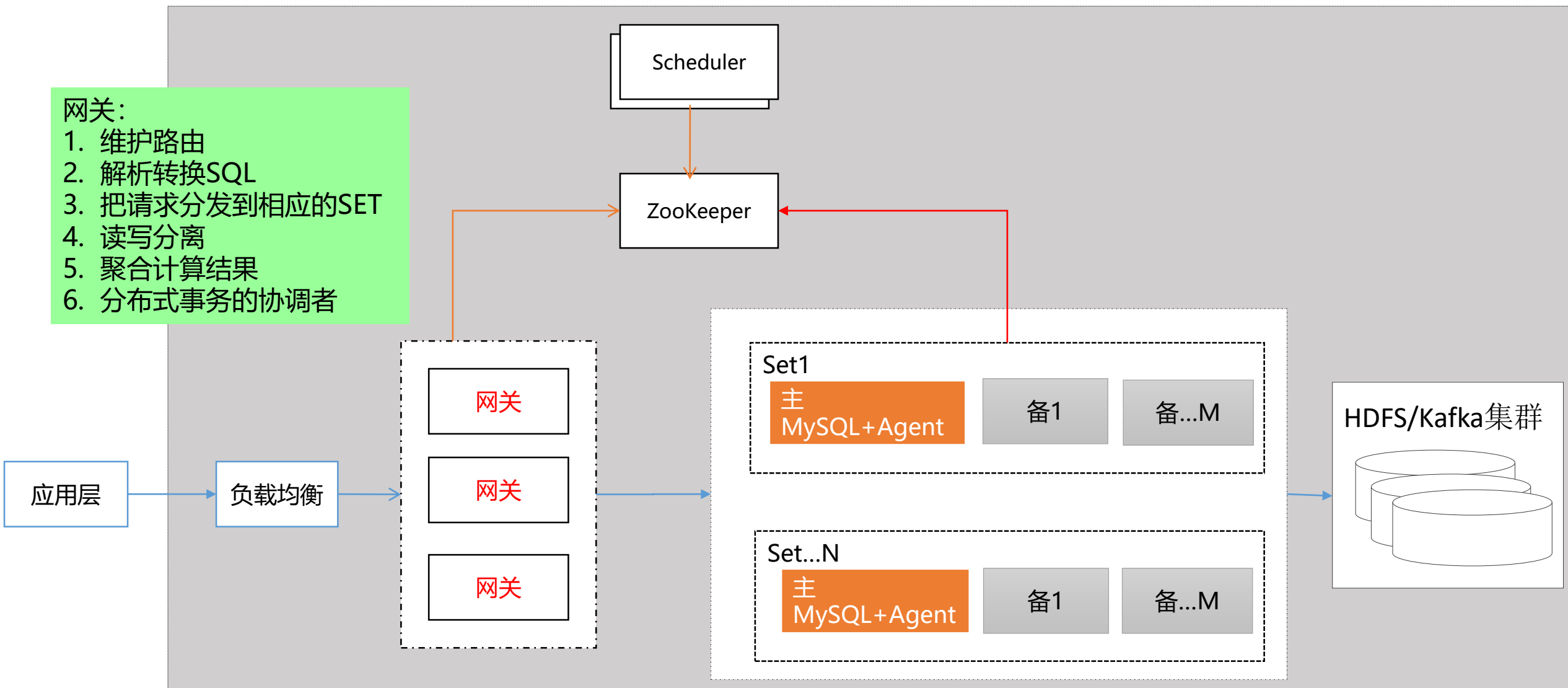
# 2 核心特性

高一致性、高可用性、数据保护  
水平扩展、高性能、自动化运营体系

# TDSQL核心架构

网关：

1. 维护路由
2. 解析转换SQL
3. 把请求分发到相应的SET
4. 读写分离
5. 聚合计算结果
6. 分布式事务的协调者





# 数据库引擎



默认支持

MariaDB

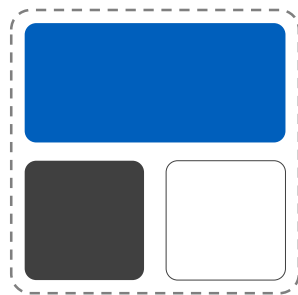
10.1

10.0

兼容MySQL 5.6

Percona/MySQL

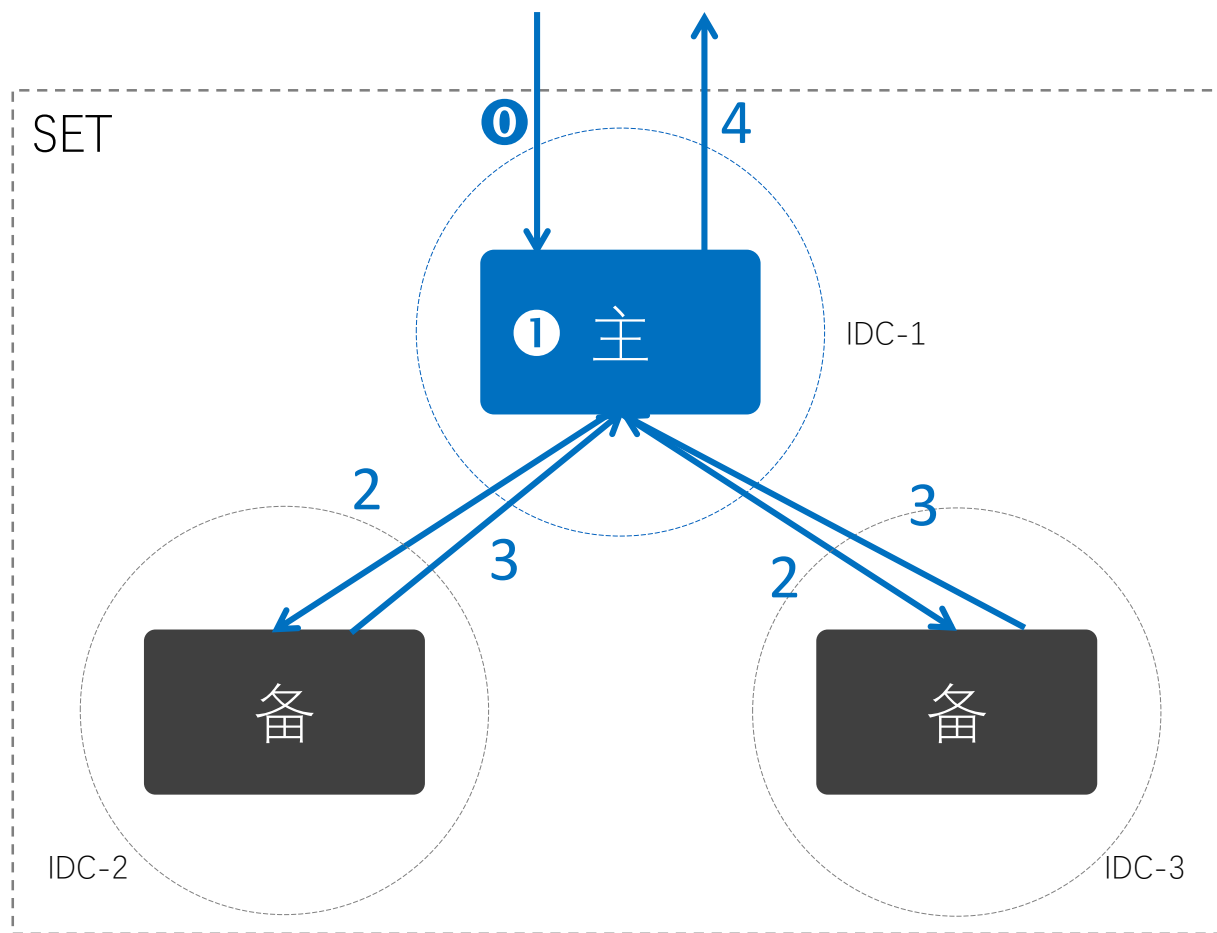
5.7



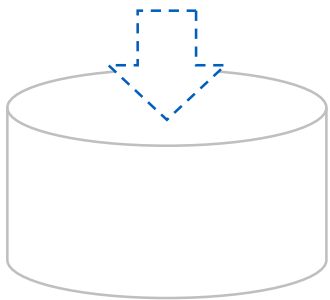
复制

Replica

# 强同步一致性协议

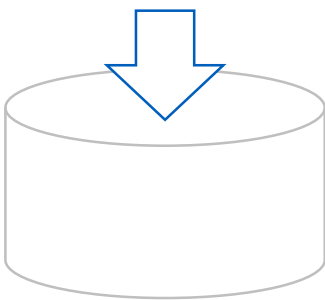


# 主备数据复制方式



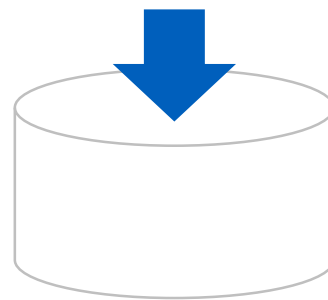
异步复制

Async replication



半同步复制

Semi-Sync replication



强同步复制

Sync replication

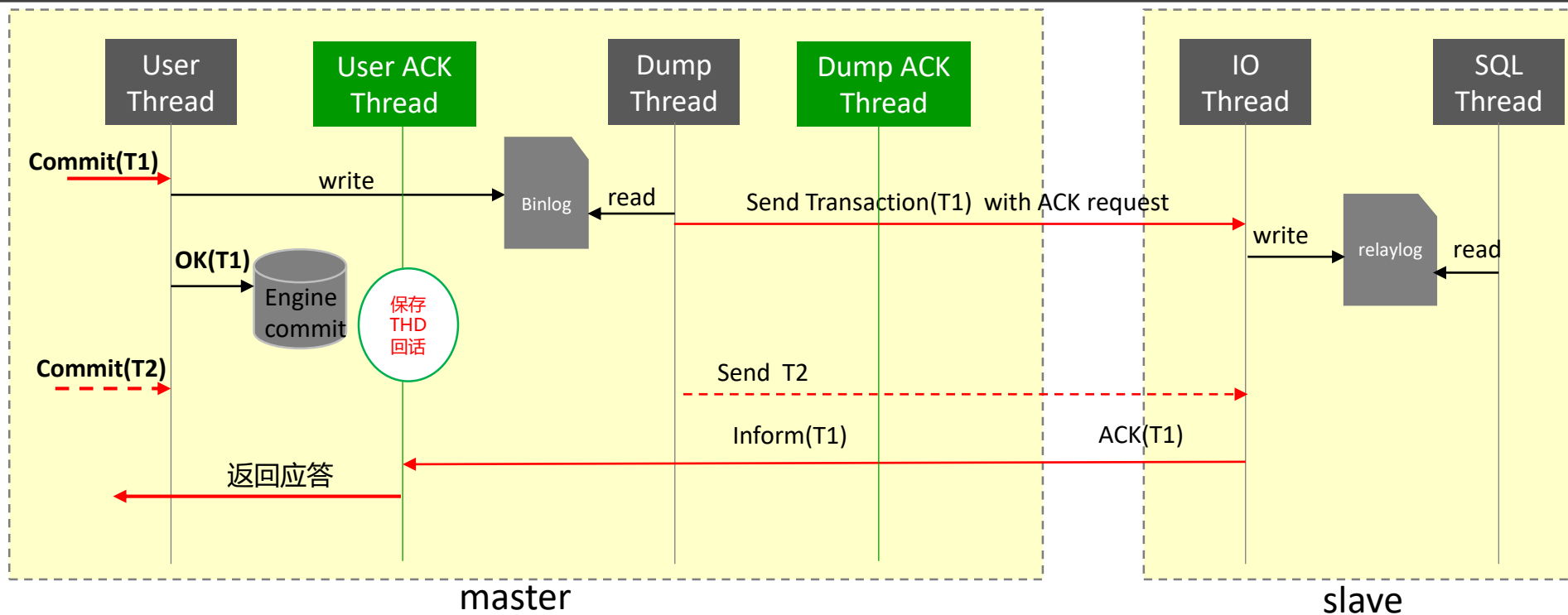
# 半同步复制的不足

1. 超时后蜕化成异步，金融场景不合适
2. 跨IDC的情况下性能不乐观(2-3ms)

update\_non\_index.lua

复制性能对比 (跨IDC)	TPS
异步	60,000
5.6半同步	5000
5.7半同步	30000

# 用户线程异步化

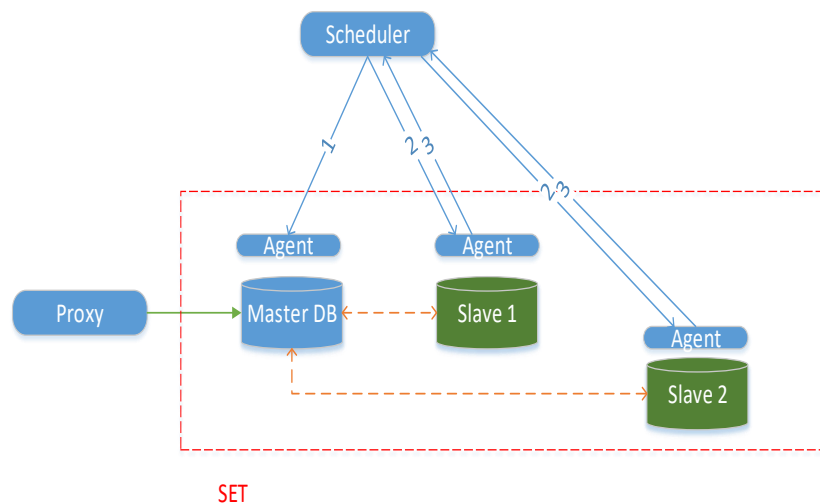


复制性能对比 (跨IDC)	TPS
异步	60,000
5.6半同步	5000
5.7半同步	30000
TDSQL强同步	60000

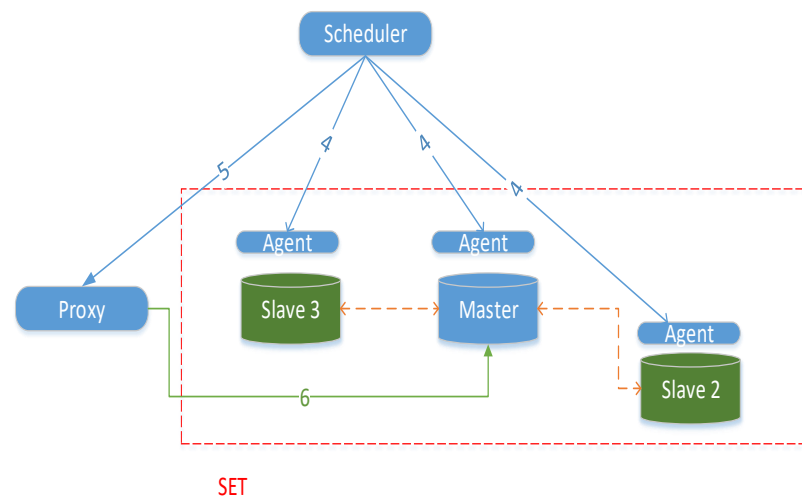
# 高一致性容灾 — 如何保证没有脏数据

## 原则：

- 1、主机可读可写，备机只读，备机可以开放给业务查询使用
- 2、任何时刻同一个SET不能有两个主机

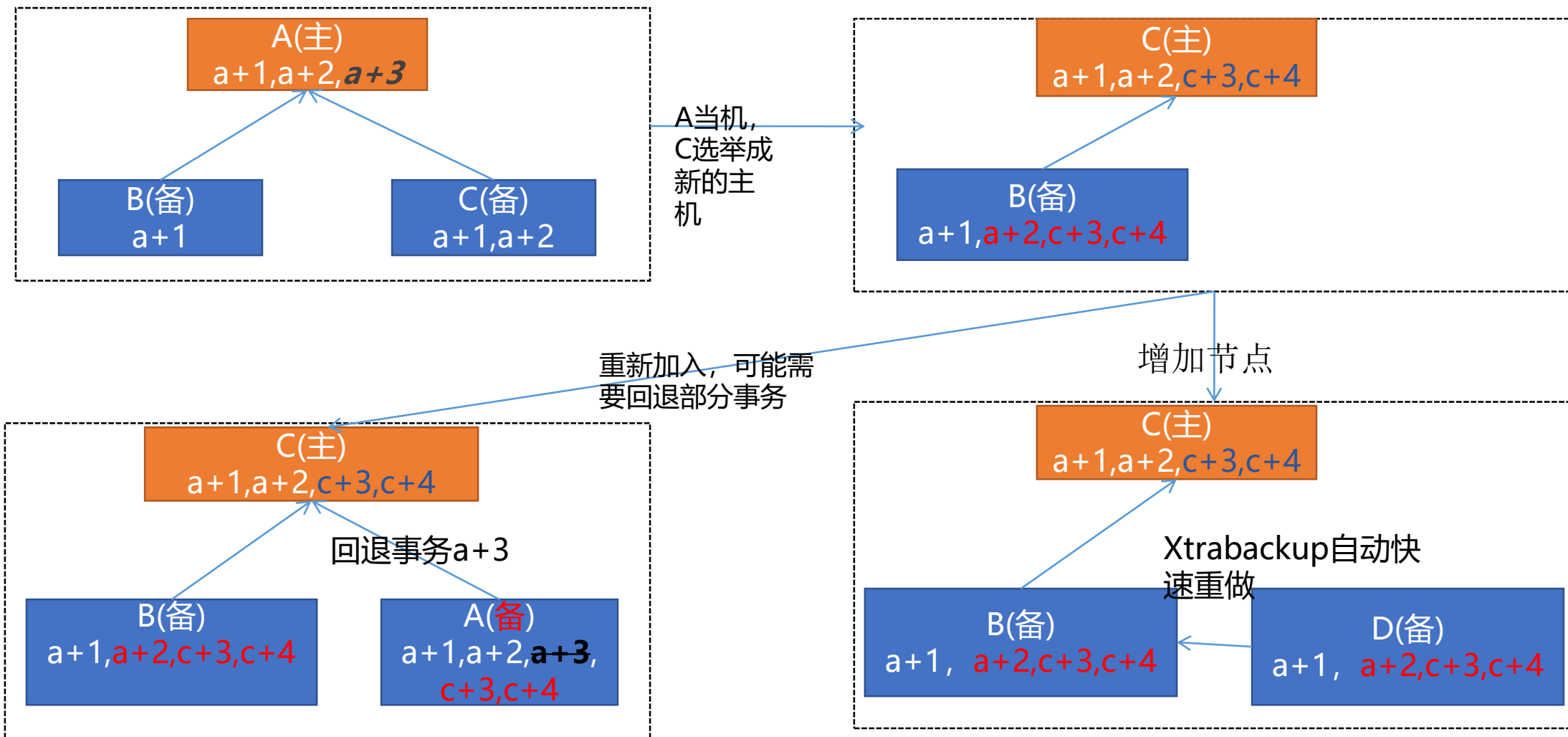


- 1、主DB降级为备机
- 2、参与选举的备机上报最新的binlog点
- 3、scheduler收到binlog点之后，选择出binlog最大的节点



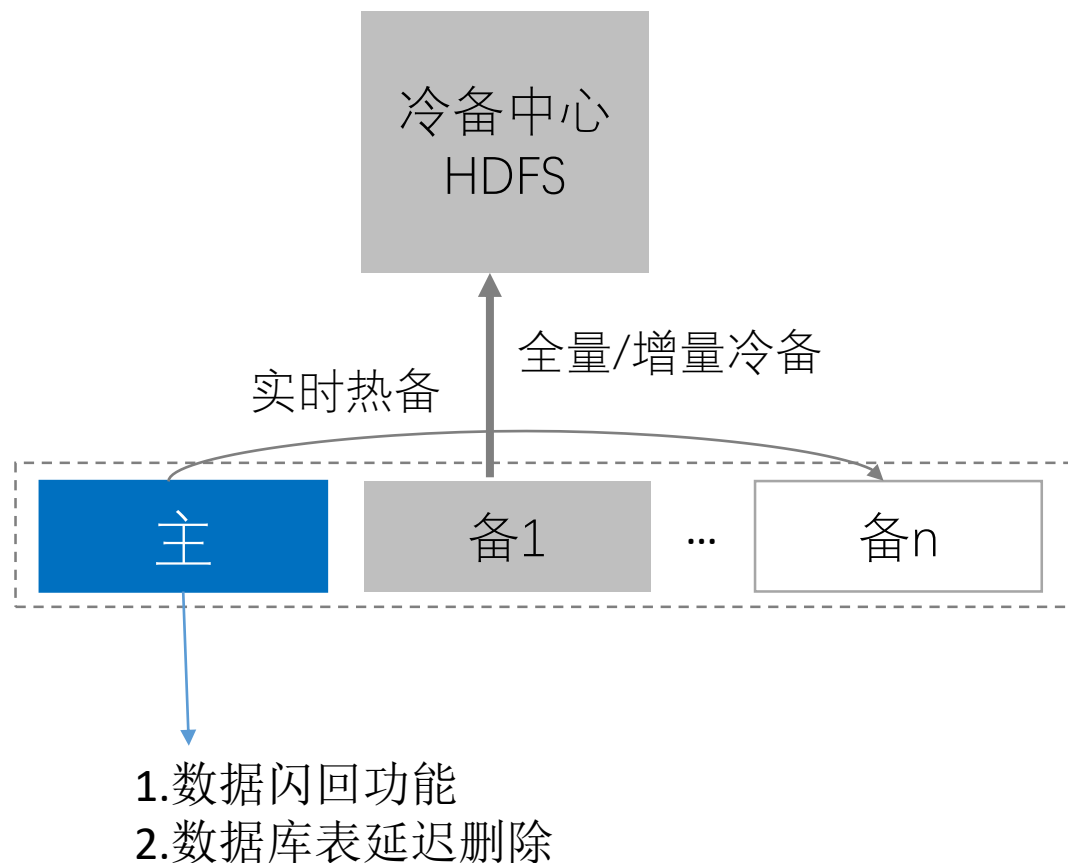
- 4、重建主备关系
- 5、修改路由
- 6、请求发给新的主机

# 数据高可用性的保障机制（恢复）





# 数据备份和恢复



## 物理备份

1. 利用xtrabackup备份物理文件。
2. 每天凌晨或者指定时间点备份一次。
3. 压缩备份到HDFS。
4. 支持增量备份
5. 支持通过管理平台直接触发备份

## 逻辑备份

1. 利用mydumper备份工具。
2. 每天备份一次。
3. 压缩备份到HDFS。
4. 每个库、表结构、表数据独立备份，恢复时可以单独恢复。
5. 支持通过管理平台直接触发备份

## binlog备份

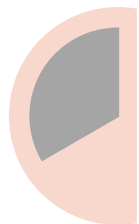
1. 实时备份binlog到hdfs。lz4压缩。

# 全面安全防护

前期

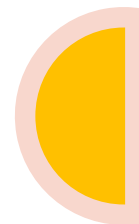


- IP白名单
- SSL连接加密
- SQL防火墙



中期

- 文件透明加密
- 网络隔离
- 运维安全保障



后期

- 数据库审计
- 操作日志审计
- 服务器审计

# 读写分离

- 基于数据库账号的读写分离
- 自动分析事务进行读写分离
- 基于Hint的读写分离

```
//主机读//  
select * from emp order by sal, deptno desc;  
//从机读//  
/*slave*/ select * from emp order by sal, deptno
```

## 只读帐号设置

×

只读帐号非全局设置，调整不会影响其他只读帐号

帐号名: onlyread

主机: %

只读请求分配策略:\*

☐ 主机 ☒ 直接报错

选择“主机”则备机不可用时读取主机，否则备机不可用直接返回失败

只读备机延迟参数:\*

10 秒

如果备机延迟超过本参数设置值，系统将认为备机发生故障  
建议该参数值大于10。

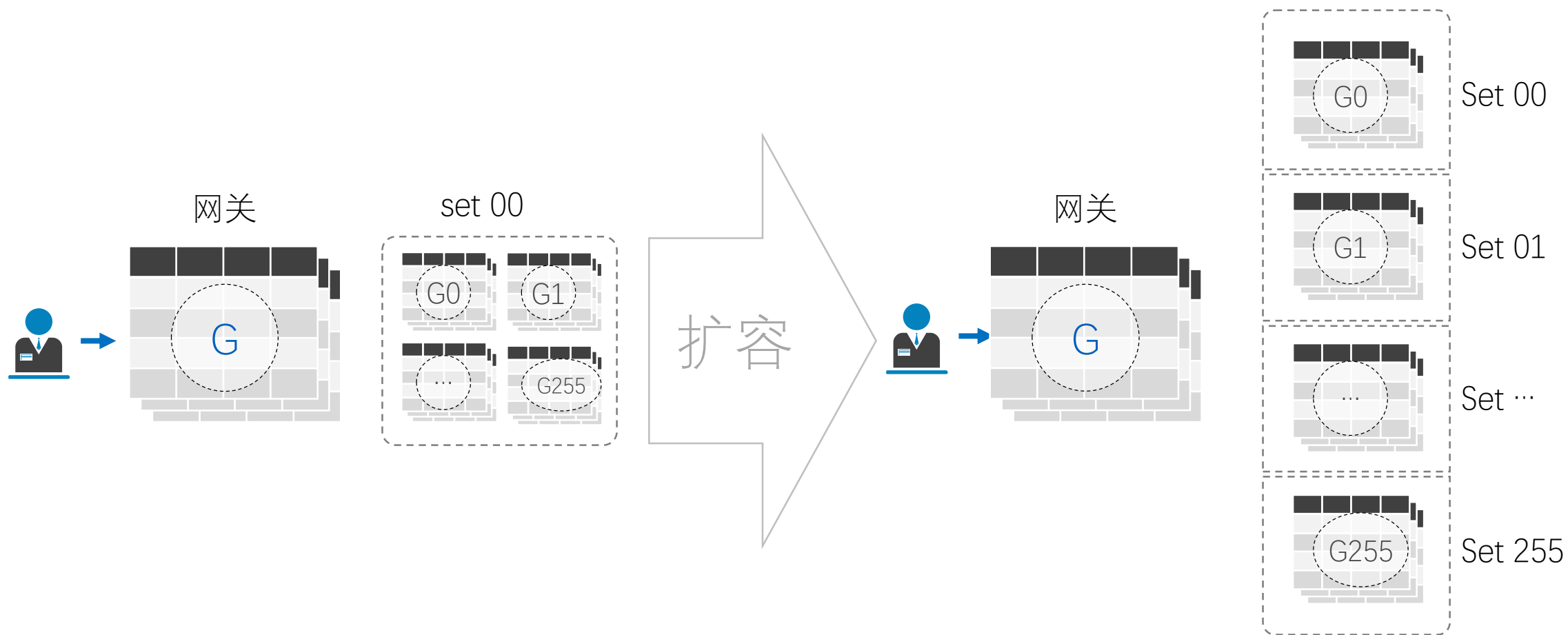
确定

取消

# 3 分布式实践

扩展性、分布式事务

# Shard水平扩容



# 三种表

- **Shard表**

- create table account( user int , payamt int, c char(20) ,PRIMARY KEY (user) ) **shardkey=user;**
- create table bill( user int , billno int, c char(20) ,PRIMARY KEY (user) ) **shardkey= user;**
- create table dummytable( seqno int , c char(20) ,PRIMARY KEY (seqno) ) **shardkey= seqno;**

- **NoShard表**, 如一些简单的配置表

- create table noshard\_table ( a int, b int key, PRIMARY KEY (a));

- **广播表**, 支持全局广播

- create table global\_table ( a int, b int key, PRIMARY KEY (a) ) **shardkey=noshardkey\_allset;**

# SQL支持-基本兼容所有SQL

- group by, order by
- max, min, sum, avg等聚合函数
- distinct, count
- Join
- Transaction (分布式事务)



# 分布式事务

XA



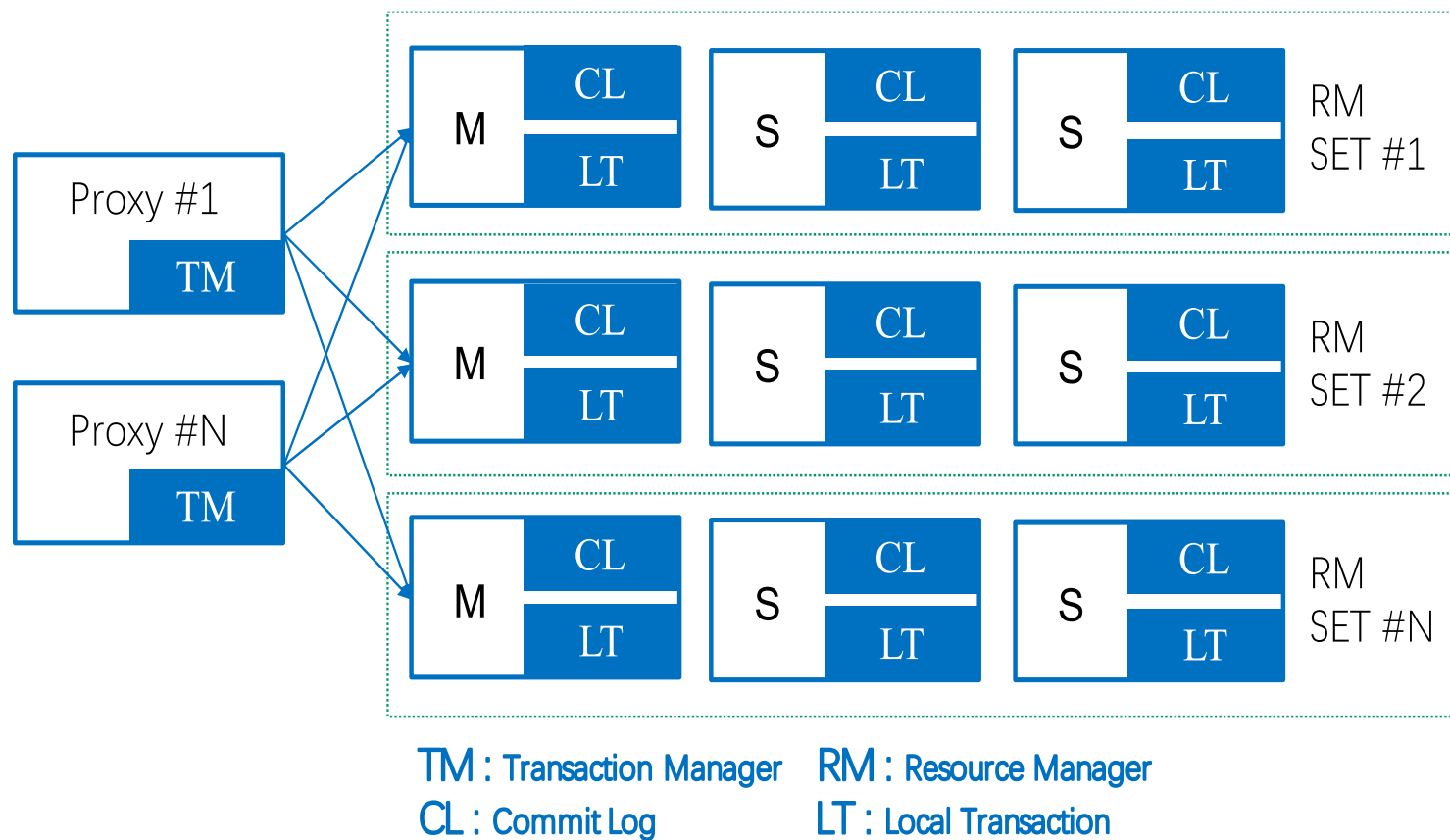
# 分布式事务

完全去中心化、性能线性增长

健壮的异常处理

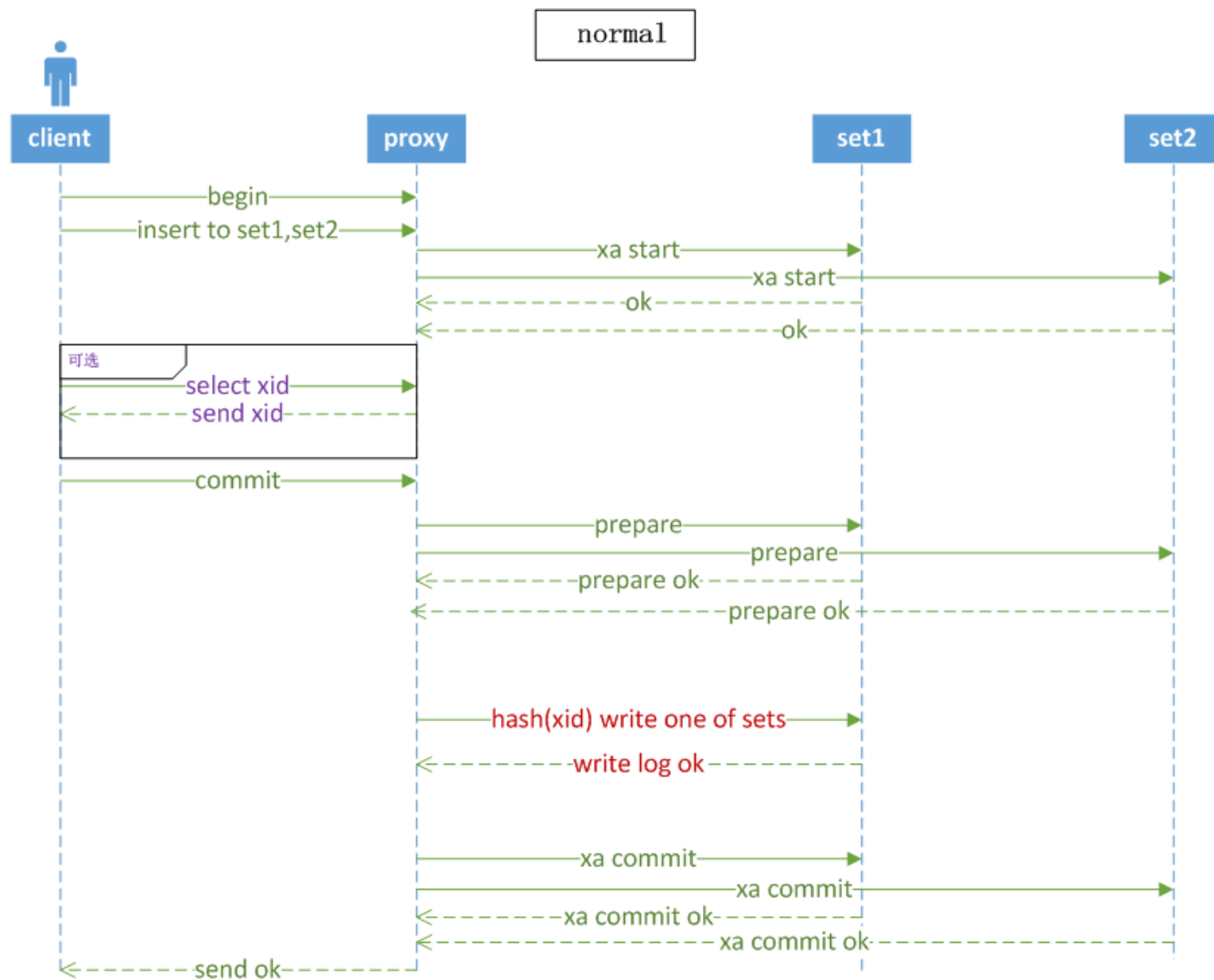
全局死锁检测机制

TPCC标准验证



# 分布式事务

- Prepare 超时或者失败
- Commit log写失败
- Commit log写超时
- Commit超时或者失败
- 异常的总结



# 选型推荐

## 中小规模

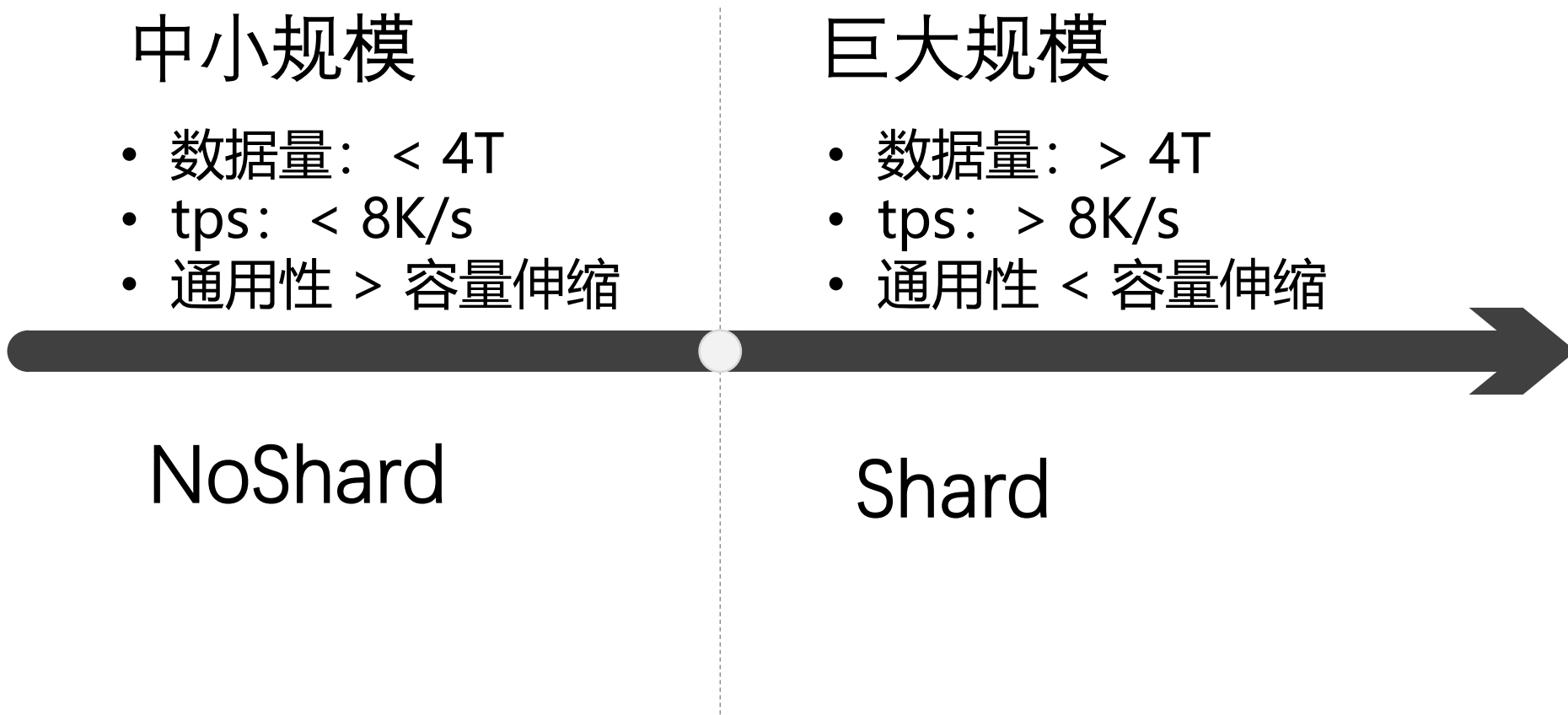
- 数据量:  $< 4T$
- tps:  $< 8K/s$
- 通用性  $>$  容量伸缩

NoShard

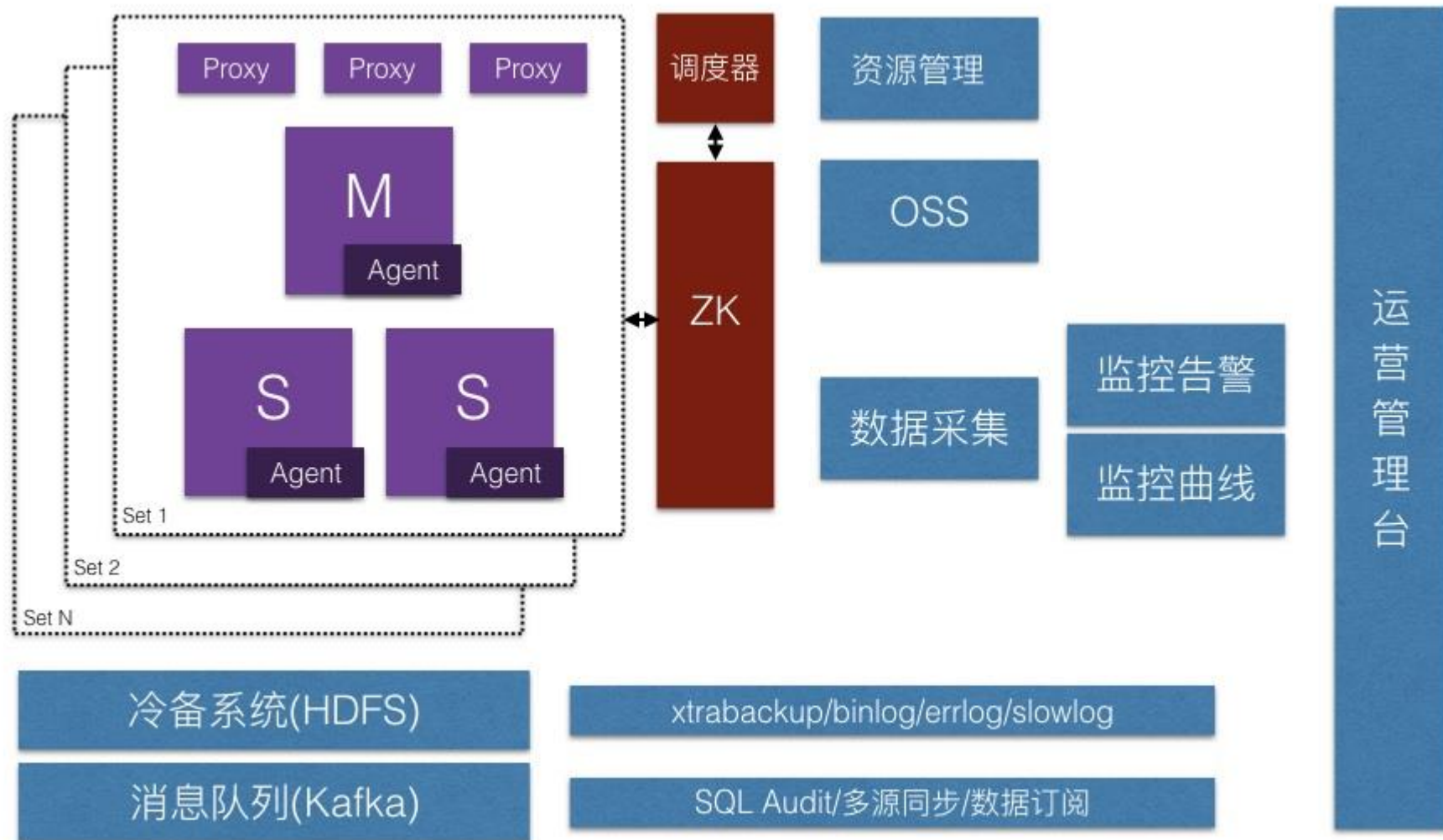
## 巨大规模

- 数据量:  $> 4T$
- tps:  $> 8K/s$
- 通用性  $<$  容量伸缩

Shard



# TDSQL整体视图



# 赤兔管理平台



MySQL指标监控详情

set\_1536733

- 实例详情
- 数据库组
- DB监控
- Proxy
- 异常
- 告警
- 日志
- 备份
- 性能

指标趋势

当前状态

总请求量 峰值 : 9992 最新值 : 0 告警策略 : [无]

日期 : 2018-12-11

Search for...

实例备份配置

物理备份 : ☒ 开启 ☐ 关闭

group\_1531727761\_858077 / 计平-上海集群 / 业务 (微信渠道订单(tinasu)) / VHOST ( )

运营状态 : ☒ 待运营 ☒ 运营中 ☐ 已下架

查询

设置【告警策略】

设置【运营状态】

设置【告警屏蔽】

	Host (IP:Port)	使用 链接数	SQL 总数	SQL 完成数	SQL 错误数	(0~5ms) 请求量	(5~20ms) 请求量	(20~30ms) 请求量	(>30ms) 请求量	存活 状态	异常 情况
<input type="checkbox"/>		324	14,860	14,861	0	14,839	22	0	0	正常	正常
<input type="checkbox"/>		328	11,948	11,950	0	11,938	11	0	1	正常	正常
<input type="checkbox"/>		124	13,646	13,646	0	13,588	58	0	0	正常	正常

binlog和冷备的保存天数 :

30

保存

模式

替换备机

删除备机

限制级别

据同步模式

# 告警

TDSQL-赤兔管理台

集群总览

实例管理

资源管理

DB汇总监控

网关汇总监控

VIP管理

调度与管理系统

跨城同步

运维操作日志

告警分析

- 告警查询
- 告警策略
- 异常查询
- 告警管理-跨集群

集群管理

告警配置云平-成都集群（合并版V12）

监控对象: ---所有--- 查询 新增策略分组+ 设置【告警推送接口】

对象	类型	命名	操作
实例	公共	实例告警策略	修改
MySQL	公共	MySQL告警策略	修改
Proxy	公共	Proxy监控策略	修改
Zookeeper	公共	ZK告警策略	修改
多源同步任务	公共	多源同步监控策略	修改
OSS任务	公共	OSS任务流程监控策略	修改
集群容量	公共	容量统计告警	修改
HDFS	公共	HDFS告警策略	修改
KAFKA	公共	Kafka告警策略	修改
Scheduler	公共	Scheduler告警策略	修改
设备资源	公共	设备资源监控策略	修改
OSS服务	公共	OSS服务监控策略	修改
Manager	公共	Manager告警策略	修改

# 扁鹊-TDSQL数据库诊断

锁等待诊断 (lock\_wait) 分数: -10

长会话诊断 (long\_session)

长会话 (long\_session)

ID	用户
159363	test
159364	test

慢查询分析 (slow\_sq

TOP时耗慢查询 分数:

用户(user@host)	se(S
test	[主
	[主
	[主
	[主
	[主
	[主
	[主

DB状态检查 (instanc

组(group)
InnoDB Cache

实例详情

数据库管理

DB监控

Proxy监控

异常会话

实例监控

告警查询

日志管理

备份&恢复

数据迁移

性能分析

检测报告

SQL优化

实时诊断

表空间分布

会话检查

选择Set

set\_1538046401\_4946

选择数据库

SQL语句

SQL诊断

创建时间

诊断数据库

诊断SQL

WHERE id=5006

WHERE id=5013

example_sql)	建议(example_sql_advice)
JM(K) FROM sbtest1 WHERE id 4991 AND 4991+99	
FROM sbtest1 WHERE id 4623 AND 4623+99	
FROM sbtest1 WHERE id 4990 AND 4990+99 ORDER BY c	建议添加索引: alter table sbtest1 add index idx_c(c)
ISTINCT c FROM sbtest1 WHERE IN 5027 AND 5027+99 ORDER BY	建议添加索引: alter table sbtest1 add index idx_c(c)
FROM sbtest1 WHERE id=4961	

	影响分数
	0

建议

未提交事务会导致后续相关事务陷入锁等待, 建议根据业务情况kill会话 2149

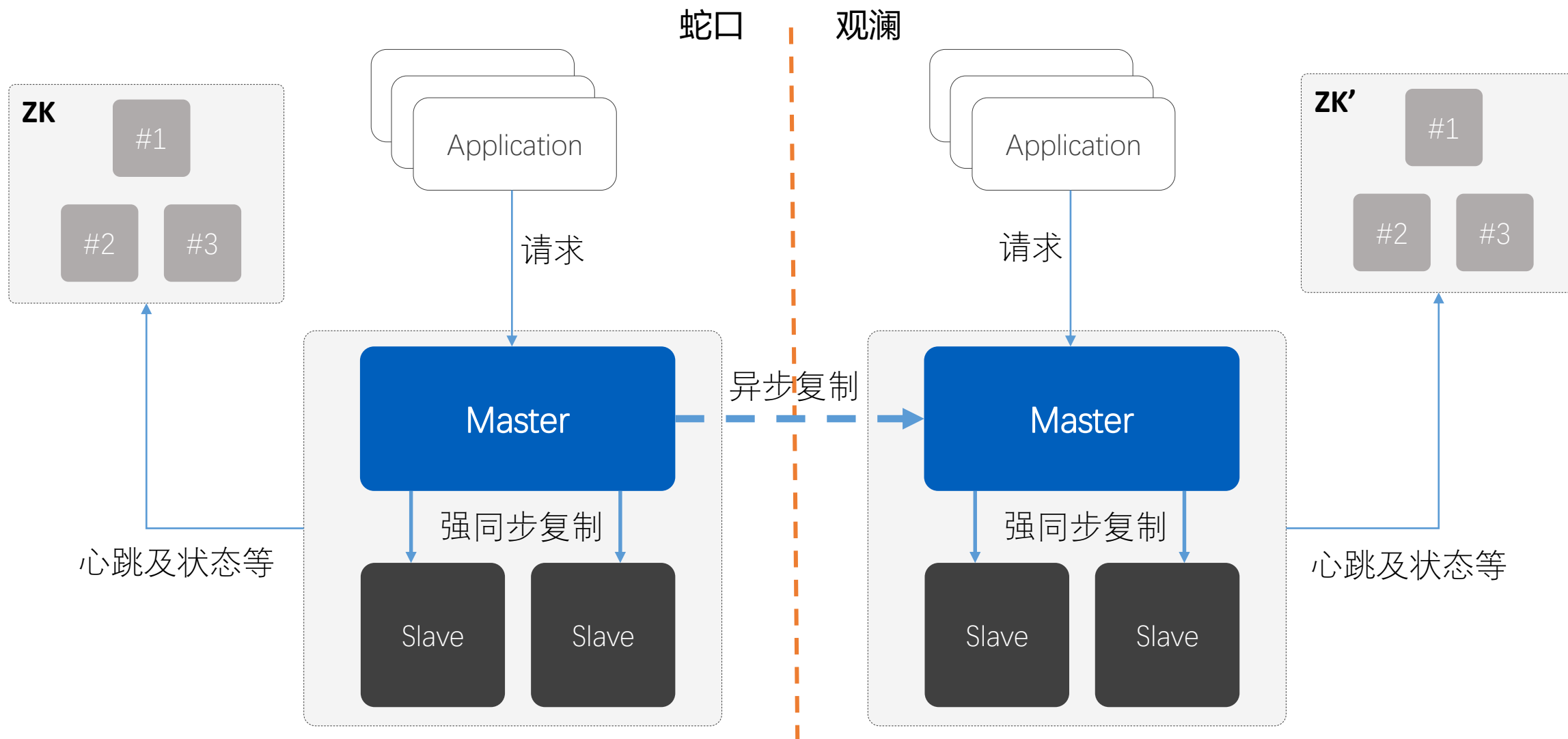
事务id: 1115581



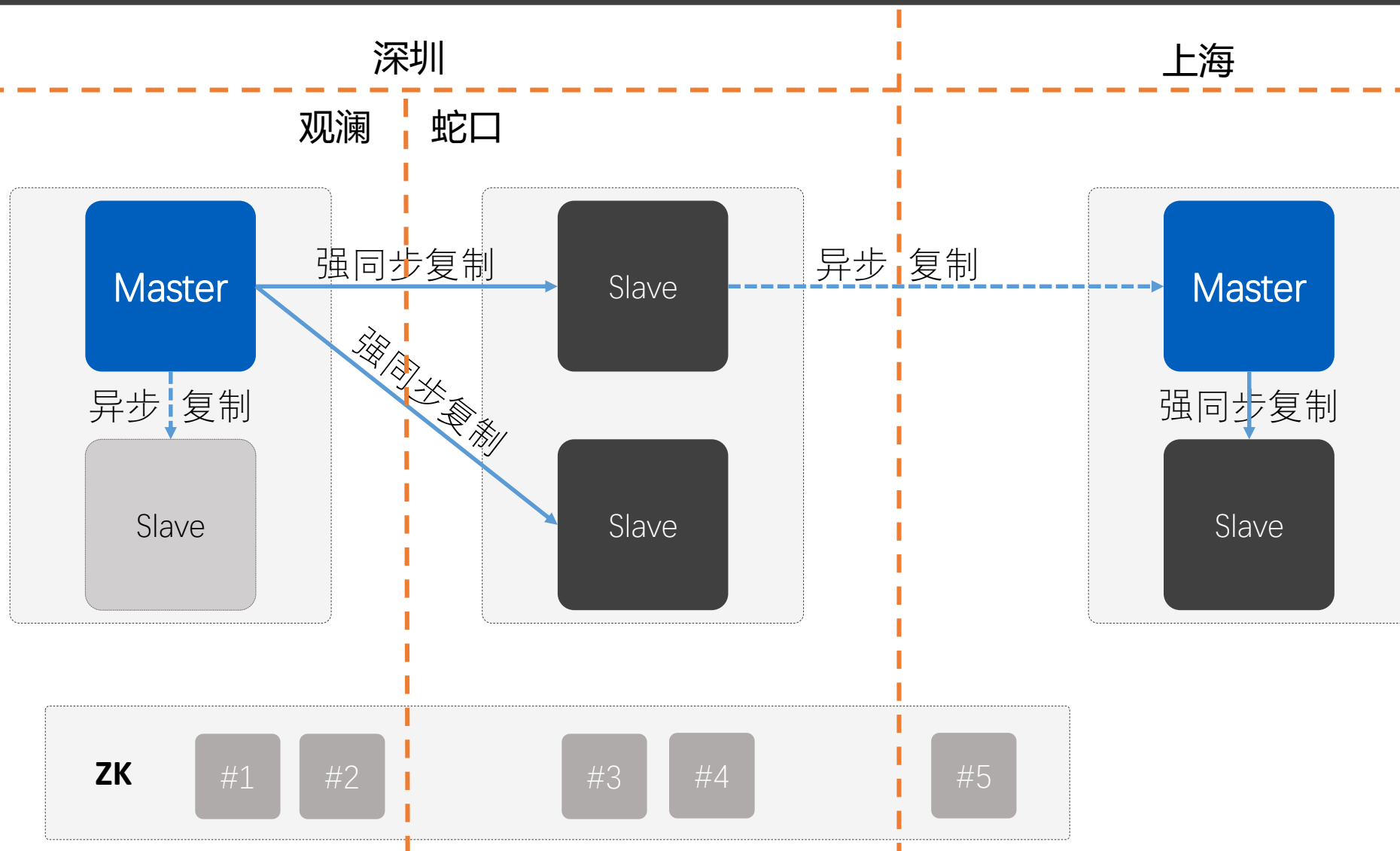
# 4 部署实践

多地多中心、强同步异步灵活部署

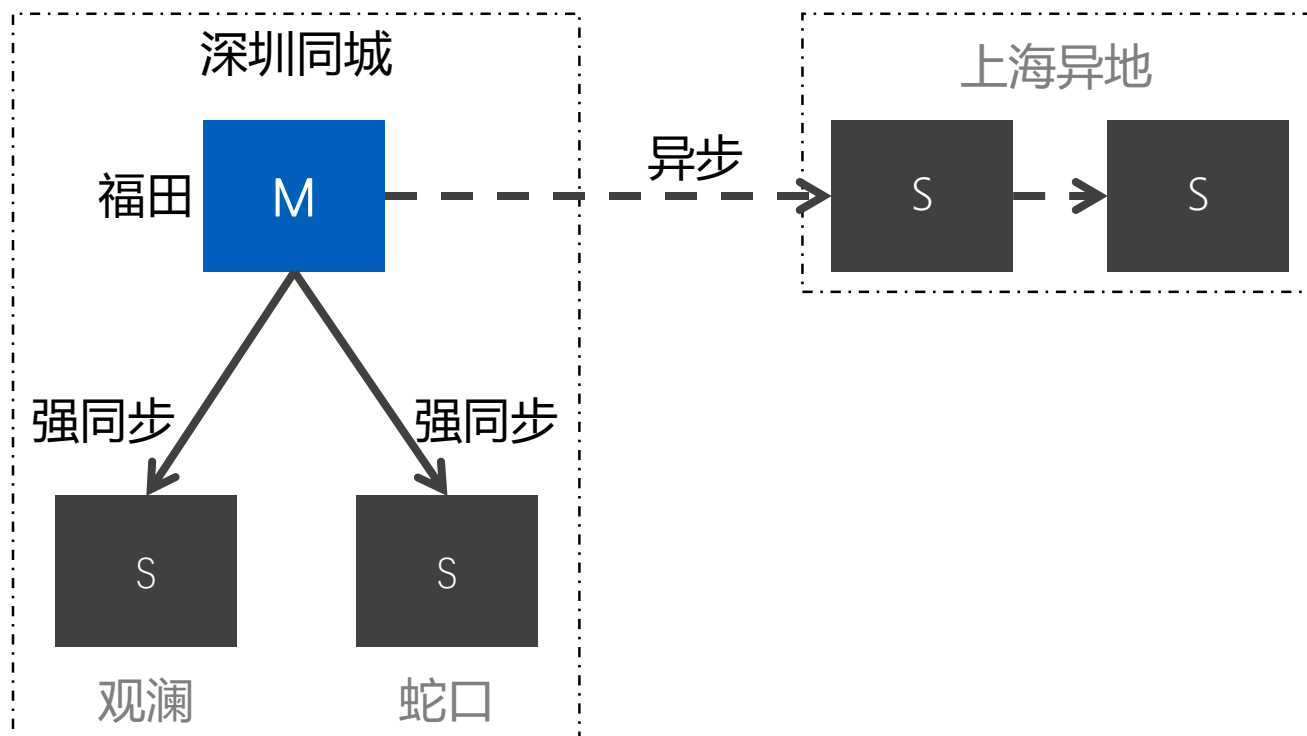
# 同城主从双中心



# 两地三中心



# 两地四中心 --(自动化切换的强同步架构)



- 同城三中心集群化部署, 简化同步策略, 运营简单, 数据可用性、一致性高
- 单中心故障不影响数据服务
- 深圳生产集群三中心多活
- 整个城市故障可以人工切换



## 关于「3306π」社区

围绕MySQL核心技术，将互联网行业中最重要  
的数据化解决方案带到传统行业中  
囊括**其他开源技术**，redis、MongoDB、Hbase、  
Hadoop、ElasticSearch、Storm、Spark等  
在全面互联网化的大趋势下，将互联网新鲜的核心技  
术理念带到传统行业里，构建良好交流互动环境  
分享干货知识，即便是赞助商，也要求如此，拒绝  
放水

「3306 $\pi$ 」社区，欢迎您的加入



社区公众号



社区QQ群



*Database for your business*