



数据技术嘉年华

Data Technology Carnival

云·数据·智能 - 数聚价值智胜未来

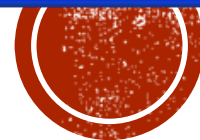
关注公众号回复help,
可获取更多经典学习
资料和文档, 电子书



腾讯金融云分布式数据库 TDSQL的架构与技术

李海翔 @那海蓝蓝

2017/11



第七届



数据技术嘉年华
Data Technology Carnival



目录

CONTENTS

TDSQL简介

TDSQL架构与分布式方案

TDSQL分布式事务处理

分布式事务处理技术



第七届



数据技术嘉年华

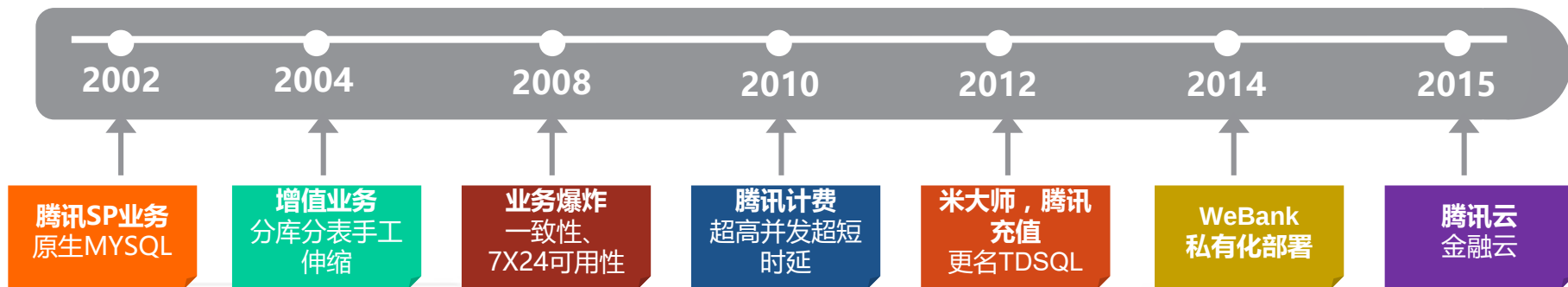
Data Technology Carnival



金融级云数据库解决方案 (CDB for TDSQL)



面向金融类业务，十年积累，亿级账户验证
腾讯公司内与计费、充值、转账、财务等核心系统90%以上都使用TDSQL！



第七届



数据技术嘉年华

Data Technology Carnival



基于MySQL生态
MySQL 100%兼容

基于OLTP场景
永不停机、高一致性
数据库集群



TDSQL 数据库的特点



跨机房部署

网络故障不影响业务

三重保障

集群内保障3套节点，单点故障整体稳定

数据强同步

主备数据完全一致

金融级安全

支持物理专享，支持数据库审计，支持加密等

可用性 : 99.999%

数据可靠性 : 99.99999%



第七届



数据技术嘉年华

Data Technology Carnival



TDSQL 数据库的特点



高一致性



高可用性



安全可靠



弹性容量



性能卓越



第七届



数据技术嘉年华
Data Technology Carnival



目录

CONTENTS

TDSQL简介

TDSQL架构与分布式方案

TDSQL分布式事务处理

分布式事务处理技术



第七届

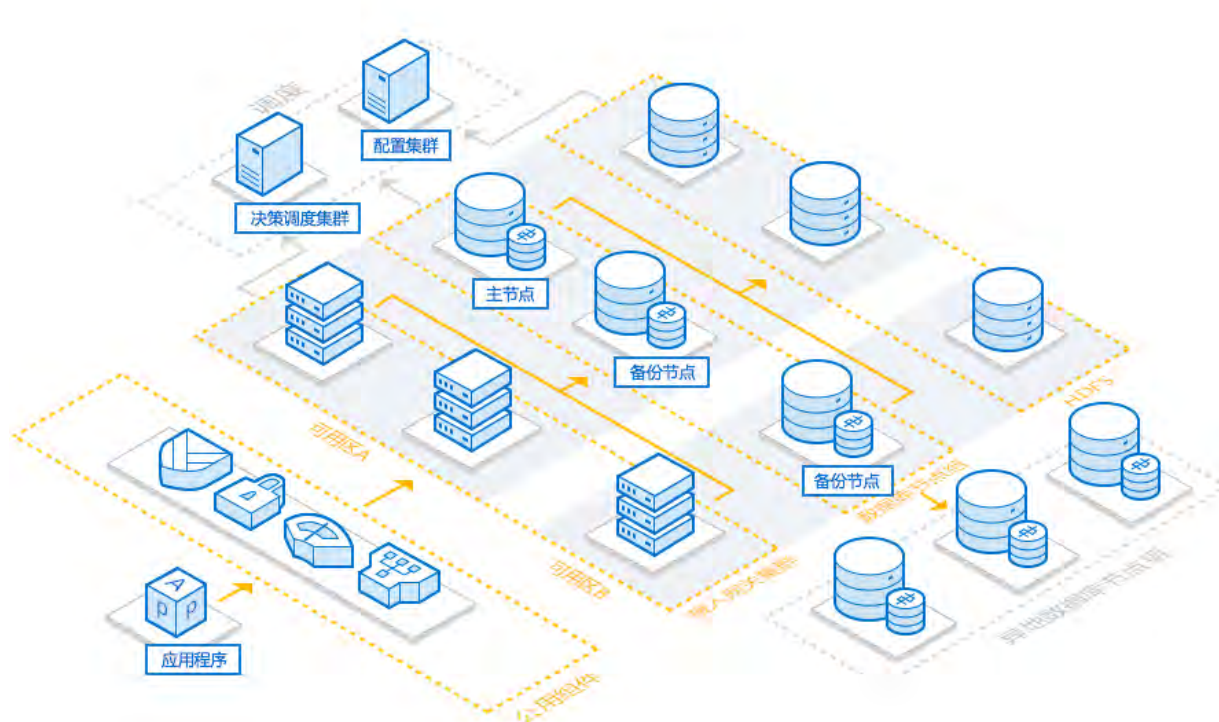


数据技术嘉年华

Data Technology Carnival



数据库部署架构



数据库节点组(SET)由MySQL数据库、监控和信息采集模块组成一主二从数据库节点。

调度集群作为集群的管理调度中心，主要管理数据库节点组、接入网关集群的正常运行

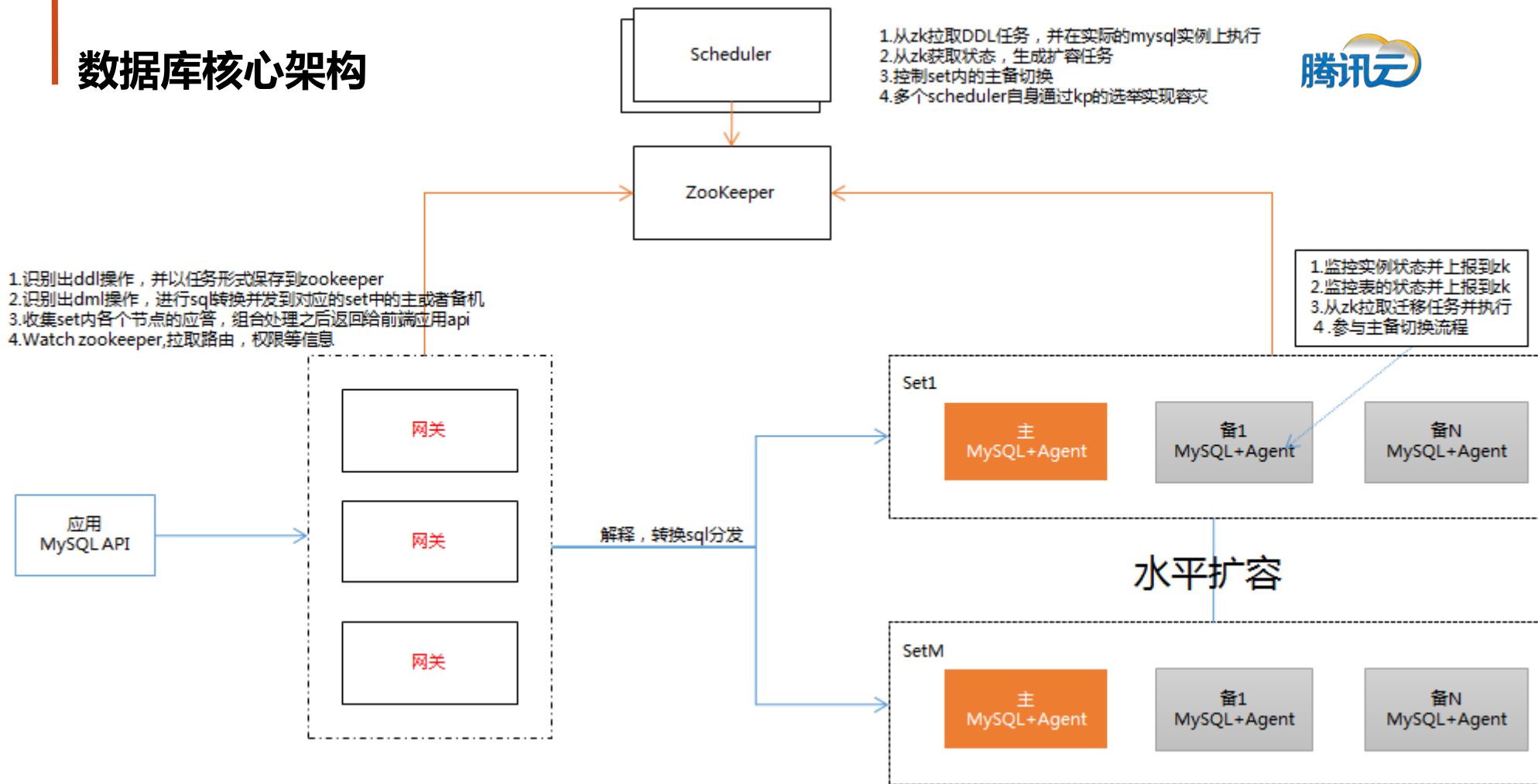
接入网关集群账号鉴权、管理连接、SQL解析、分配路由

分布式文件系统(HDFS)提供数据灾备服务，提供至少3份备份

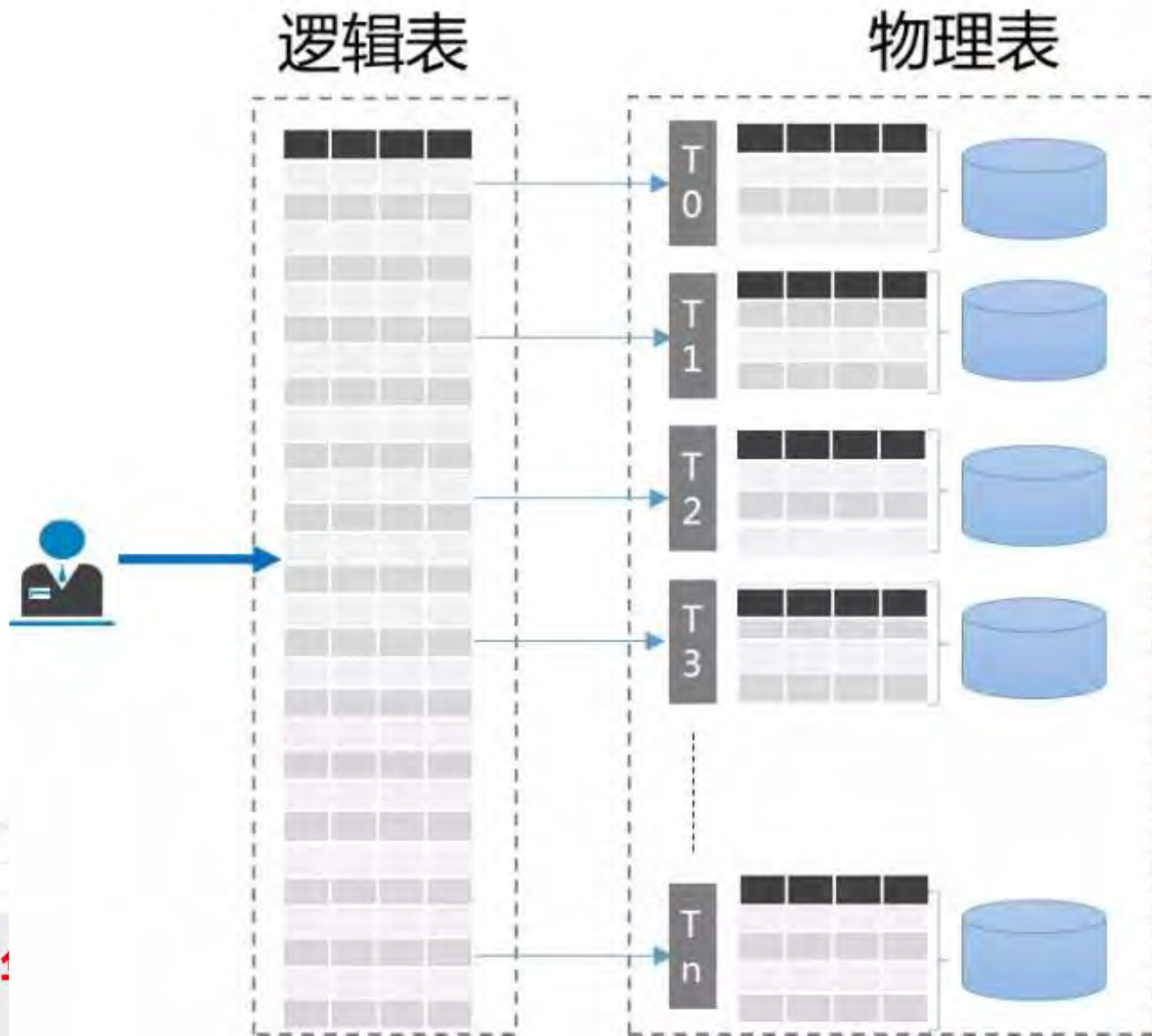
异地容灾数据库节点组部署在主节点以外的异地机房。



数据库核心架构



数据分布

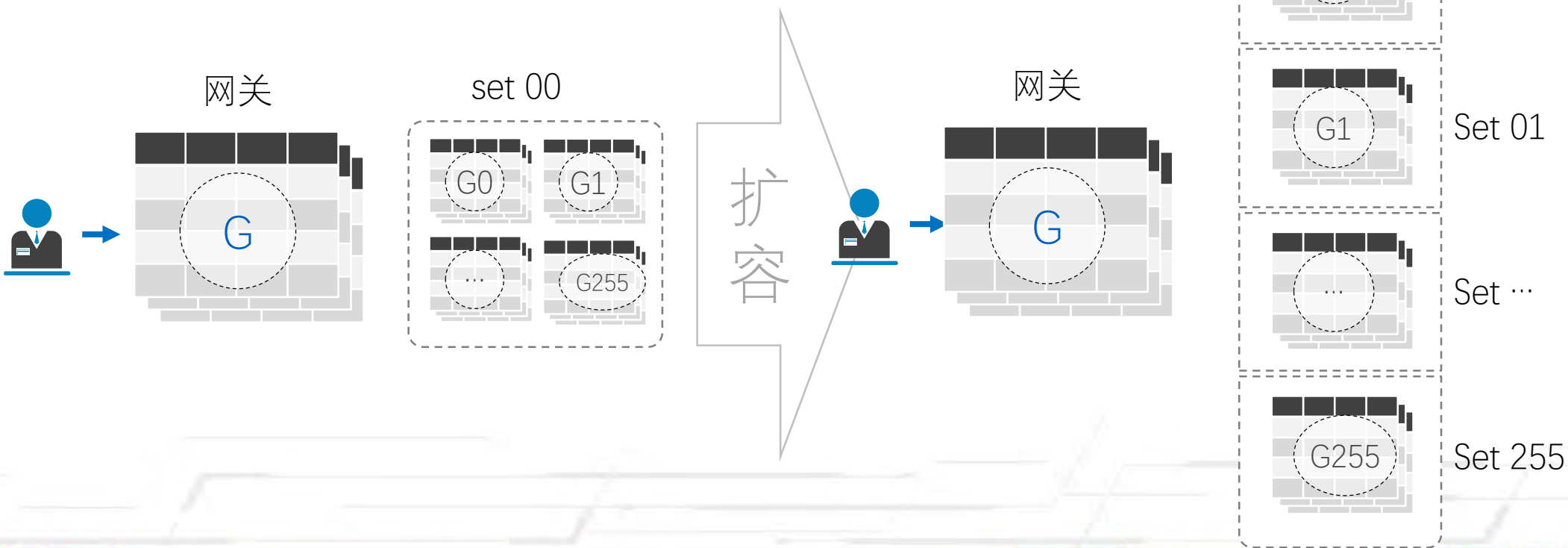


第七届



数据技术嘉年华
Data Technology Carnival

TDSQL分布式方案（自动扩容）

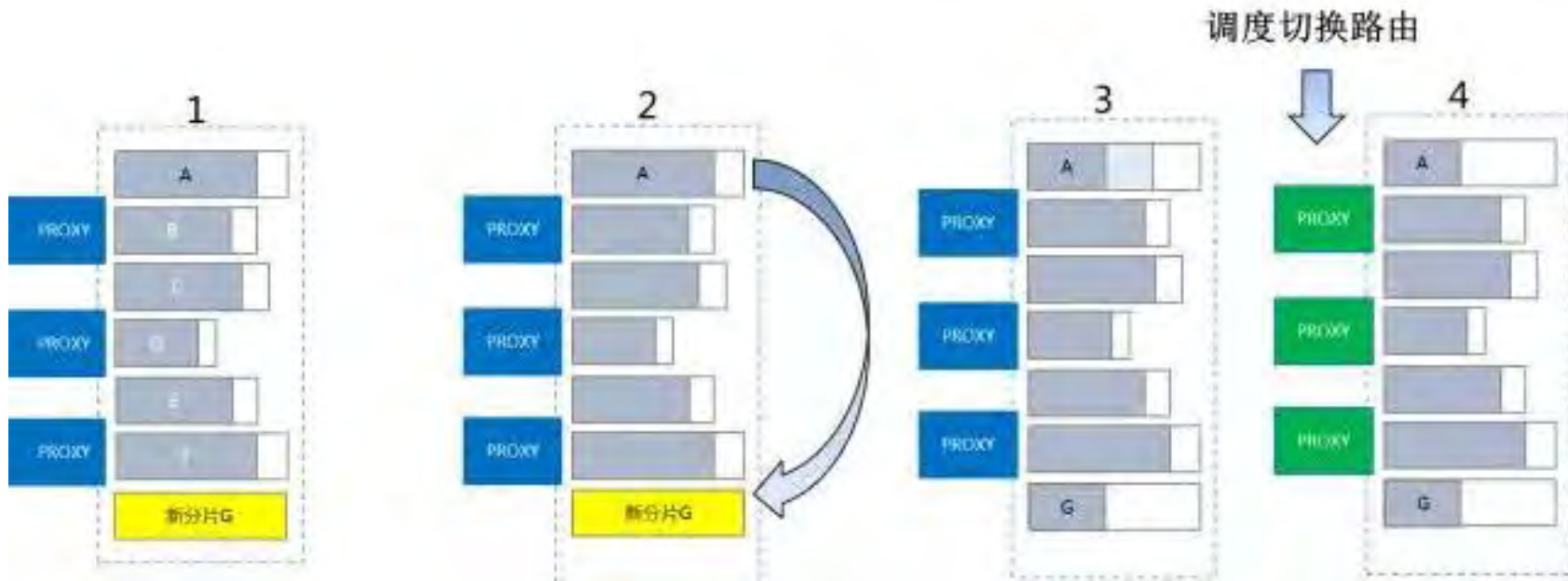


第七届



数据技术嘉年华
Data Technology Carnival





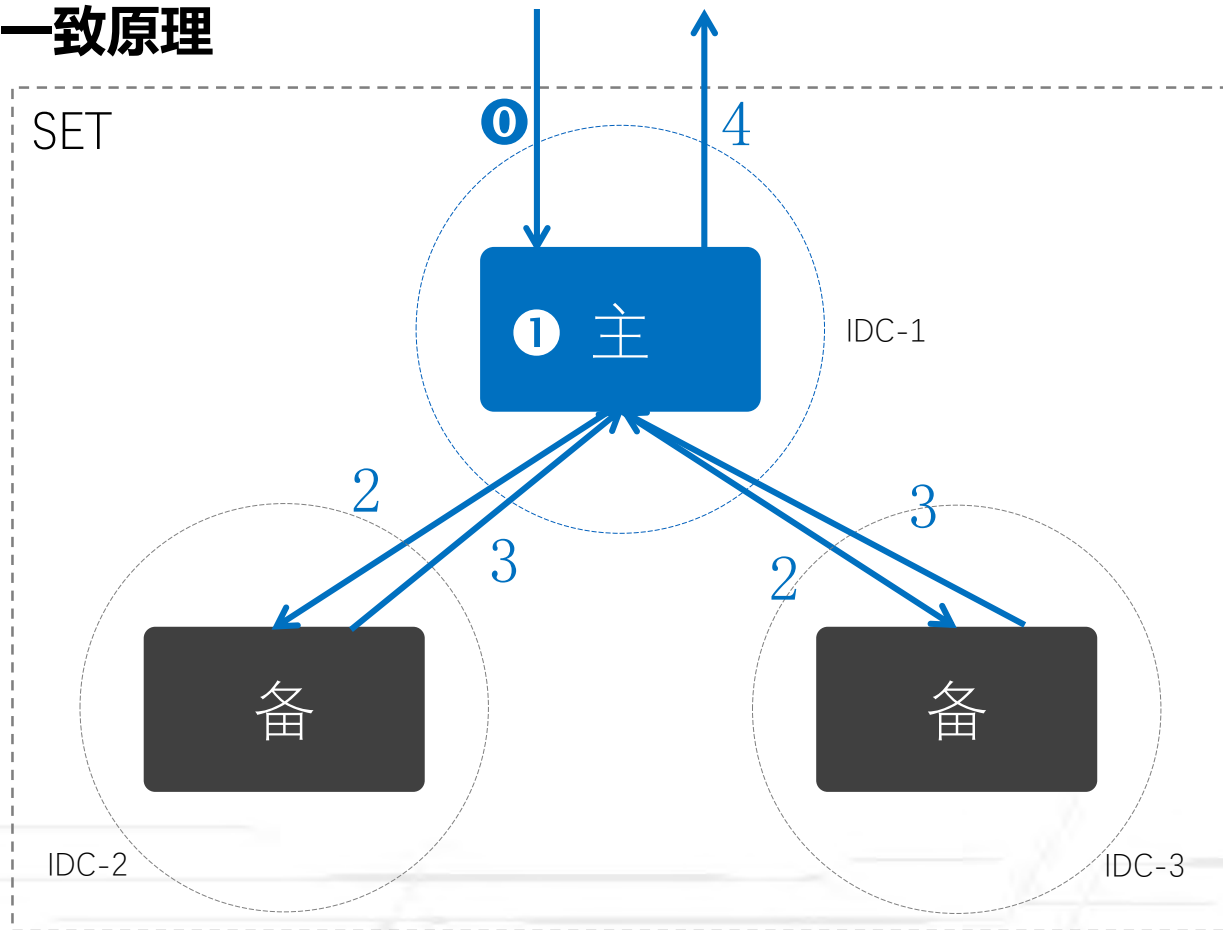
DCDB的整个迁移过程采用：

移存量数据、迁移增量数据、数据检验、再追增量、切换路由、清理
六个步骤循环迭代进行。

该能力经过腾讯内部近千个业务验证，至今未发生过一次数据丢失或错误。



TDSQL强一致原理



第七届

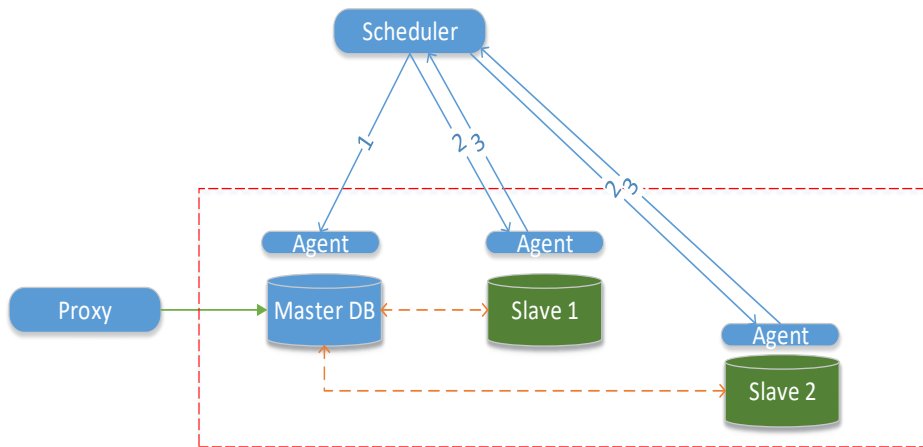


数据技术嘉年华
Data Technology Carnival

TDSQL强一致原理（确保没有脏数据）

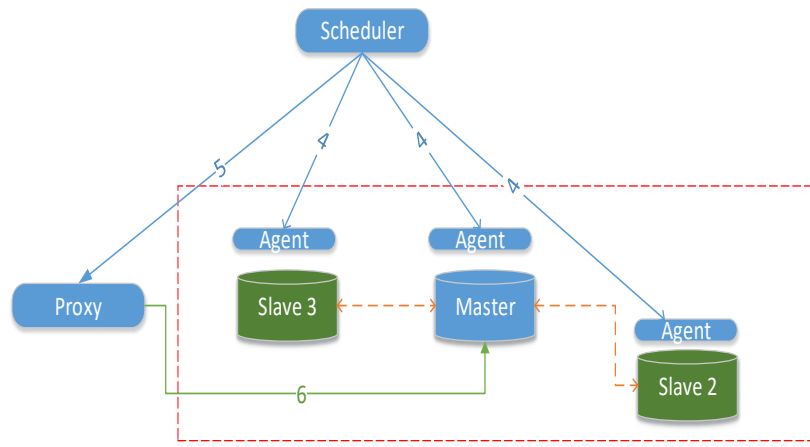


- 1、主机可读可写，备机只读，备机可以开放给业务查询使用
- 2、任何时刻同一个SET不能有两个主机
- 3、宁愿拒绝服务，不提供错误的服务，追求CAP中的C，必要的时候牺牲部分A



SET

- 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

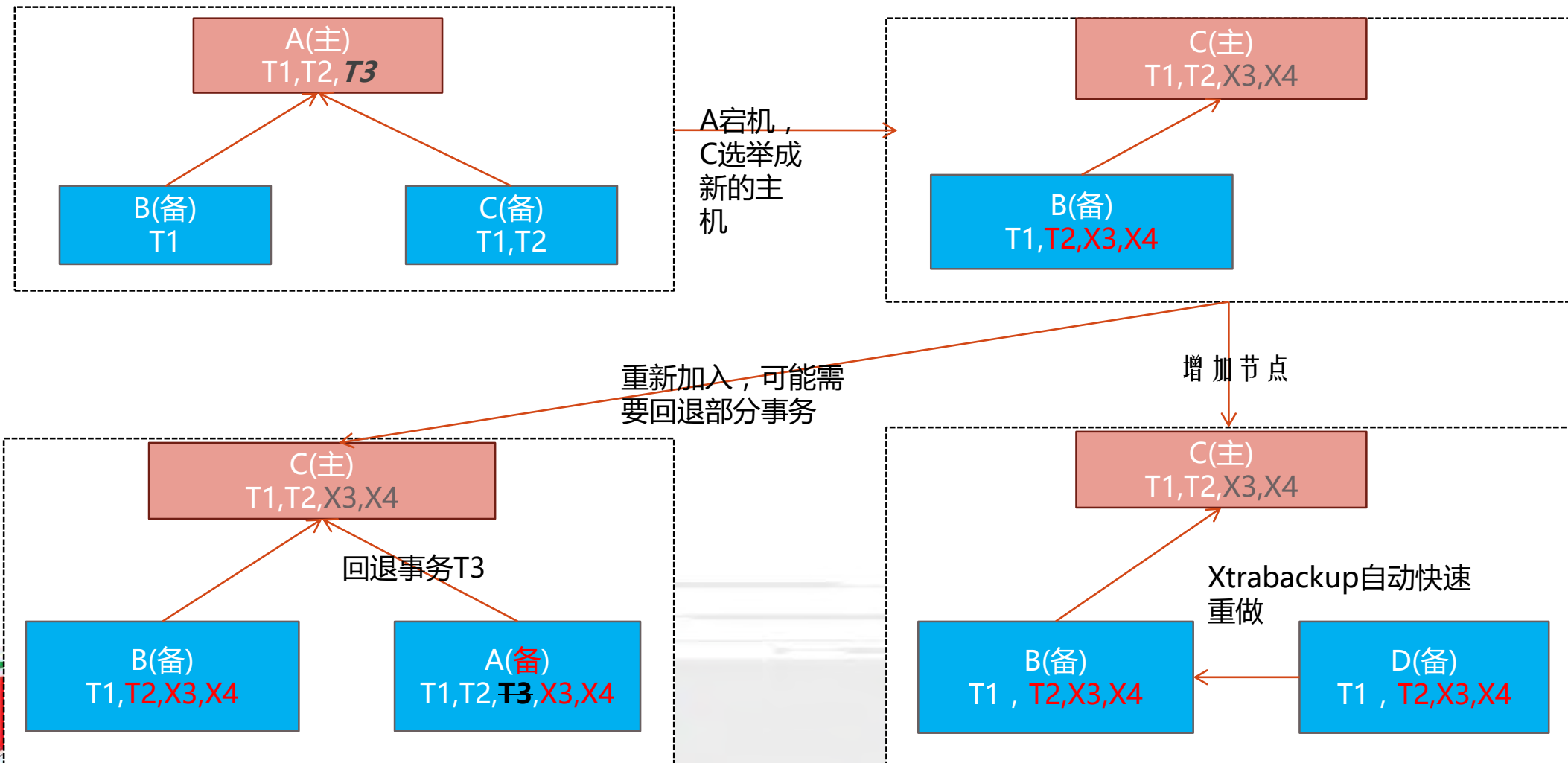


SET

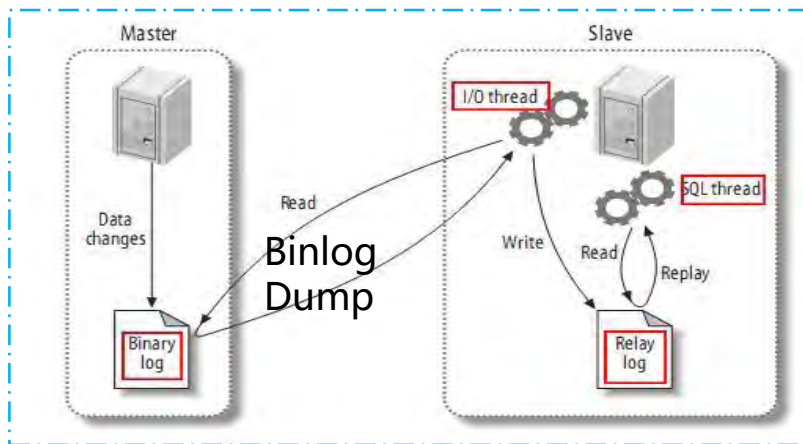
- 4、重建主备关系
- 5、修改路由
- 6、请求发给新的主机



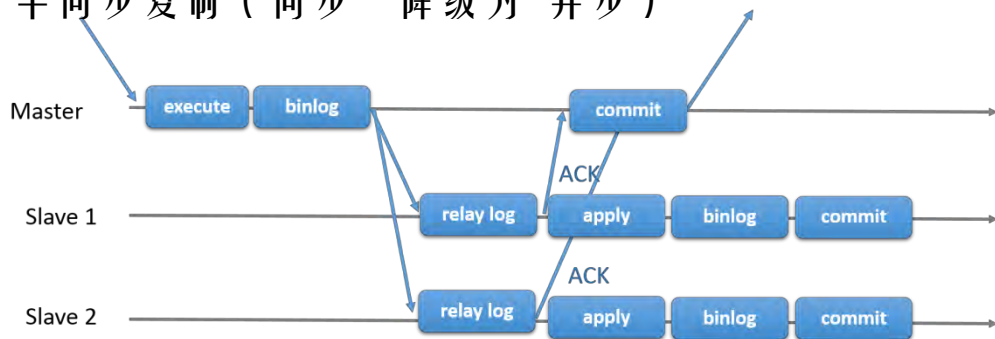
TDSQL强一致原理（恢复阶段不丢失数据）



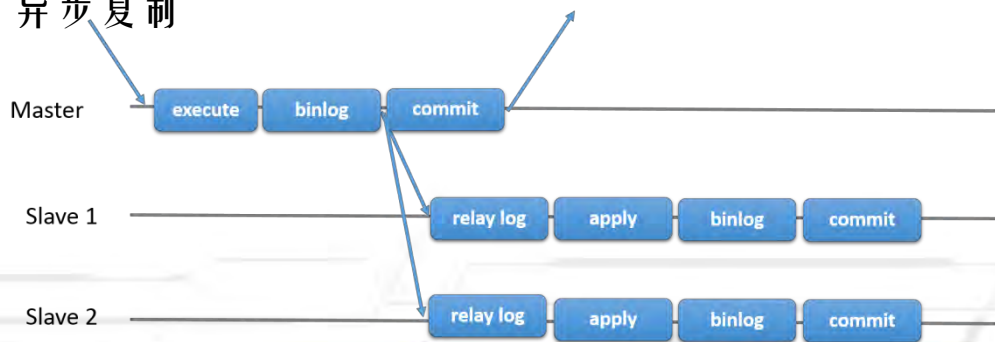
TDSQL高性能原理



半同步复制 (同步 降级为 异步)



异步复制

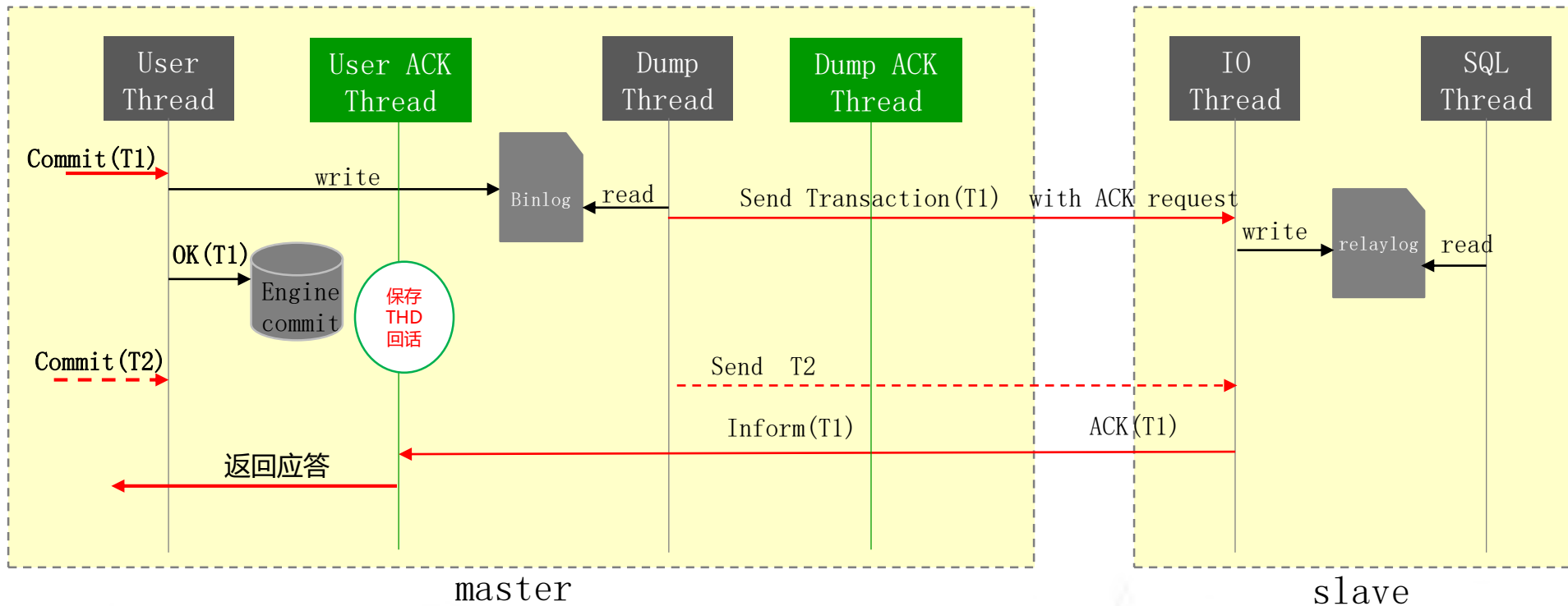


第七届



数据技术嘉年华
Data Technology Carnival

TDSQL高性能原理



主备复制方案（跨IDC）

	TPS	时耗(ms)
异步	20,000	<10
半同步	2,200	4~600ms
强同步	20000	<10
MariaDB Galera Cluster	6,000	4~10000ms



第七届



数据技术嘉年华

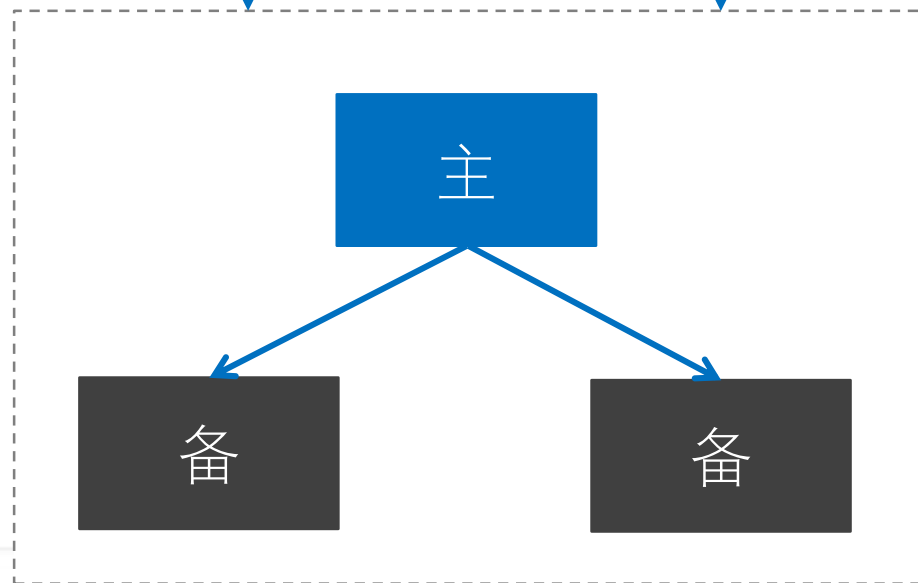
Data Technology Carnival

TDSQL高性能原理

更新索引 QPS : 10万, 99%的<10ms



纯select QPS : 50万, 99%的<5ms



环境 : ts85机型(x86, 24核(48超线程), 512G内存, 6T SSD)



第七届

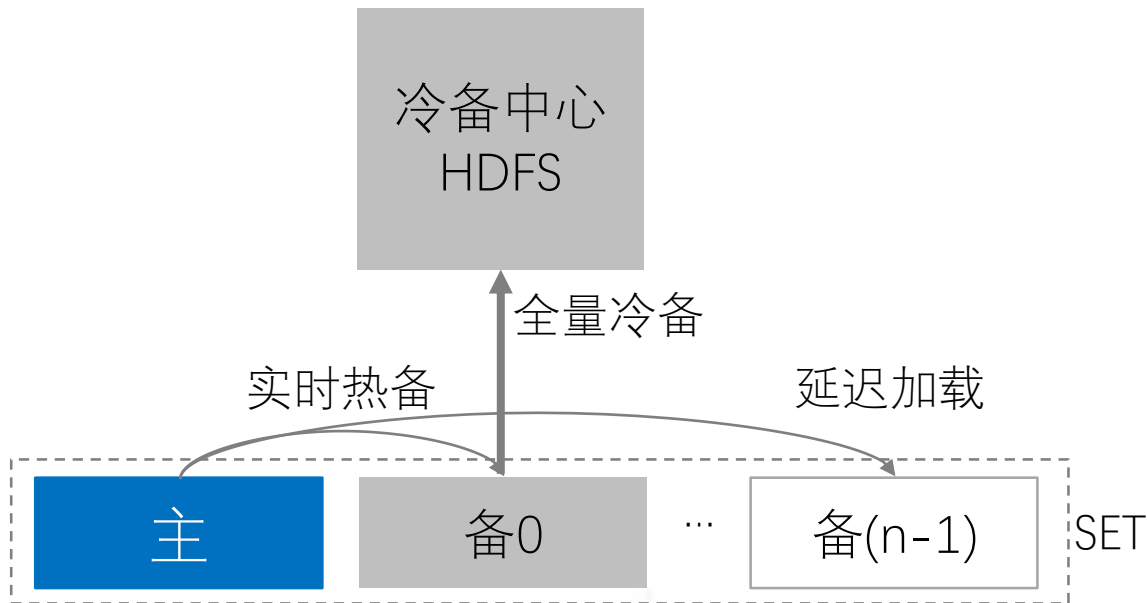


数据技术嘉年华
Data Technology Carnival

TDSQL分布式方案（可靠的备份系统）



- 数据备份
 - 热备：实时同步，实时加载
 - 冷备：快照 + binlog
- 数据恢复
 - 就地恢复（闪回/补录）
 - 新节点重建（冷备+binlog）
 - 定点回退（冷备+binlog）



第七届



数据技术嘉年华
Data Technology Carnival



- 所有的Set还是原来的NoShard实例
- 同一个用户的所有表在一起
- 小表可以广播到所有的Set
- 每个表都支持全局唯一序列号

只读帐号支持读写分离

热点更新

HASH分区

两级分区

RANGE分区

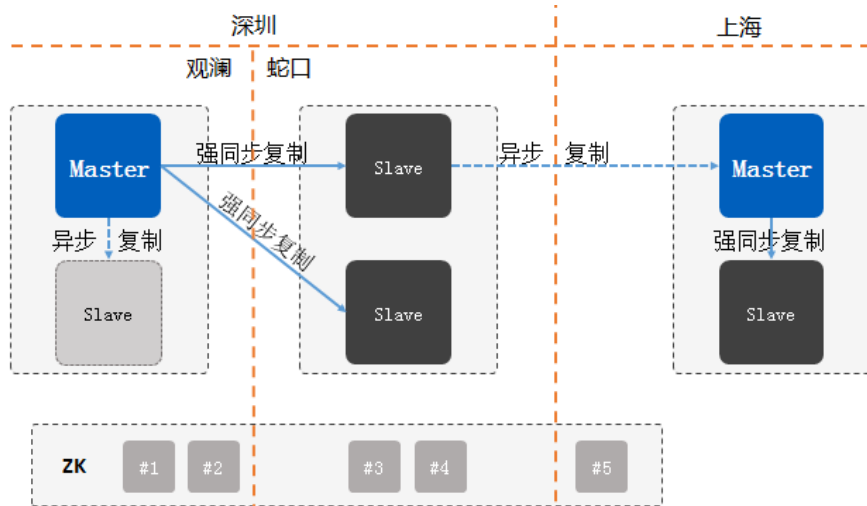
全局唯一数字序列

- group by, Order by
- Max, sum, min, ave等聚合函数
- Distinct, count(1)
- 同一个group内的join
- 事务

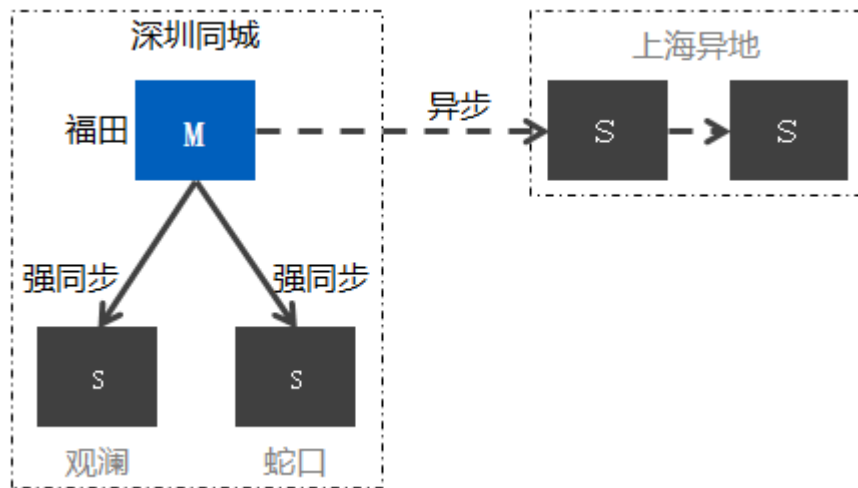
Data Buffer脏页刷出效率提升



TDSQL分布式方案（部署）



两地三中心



两地四中心 -- (自动化切换的强同步架构)



第七届



数据技术嘉年华
Data Technology Carnival



目录

CONTENTS

TDSQL简介

TDSQL架构与分布式方案

TDSQL分布式事务处理

分布式事务处理技术



第七届



数据技术嘉年华
Data Technology Carnival



TDSQL分布式事务（金融云的需求和挑战）



数据量和访问量的压力导致分库分表

- 数据量与访问负载带来的扩展需求(**scalability**) ---扩容（分库分表）
- 将数据分摊到多个**set**（存储空间和IO带宽限制）
- 将负载分摊到多个**set**（网络和CPU等资源瓶颈）
- 目标：扩容后数据库集群性能提升；
- 理想目标：扩容后性能线性提升（数据和系统耦合性导致不可能）
- **业内现状：大多只可以访问一个set --- 数据一致性的要求**



第七屆



数据技术嘉年华
Data Technology Carnival



TDSQL分布式事务（金融云的需求和挑战）



TDSQL金融业务需要分布式事务

- 同一个事务中写数据到多个set上
- 挑战：数据一致性与容灾
- 应对：定制实现应用需求比如转账
- 技术门槛非常高，大多数中小公司做不到
- 但他们是才是云计算(DaaS)的主力用户
- 应对：分布式事务
- 对用户透明，没有额外的技术门槛



第七届

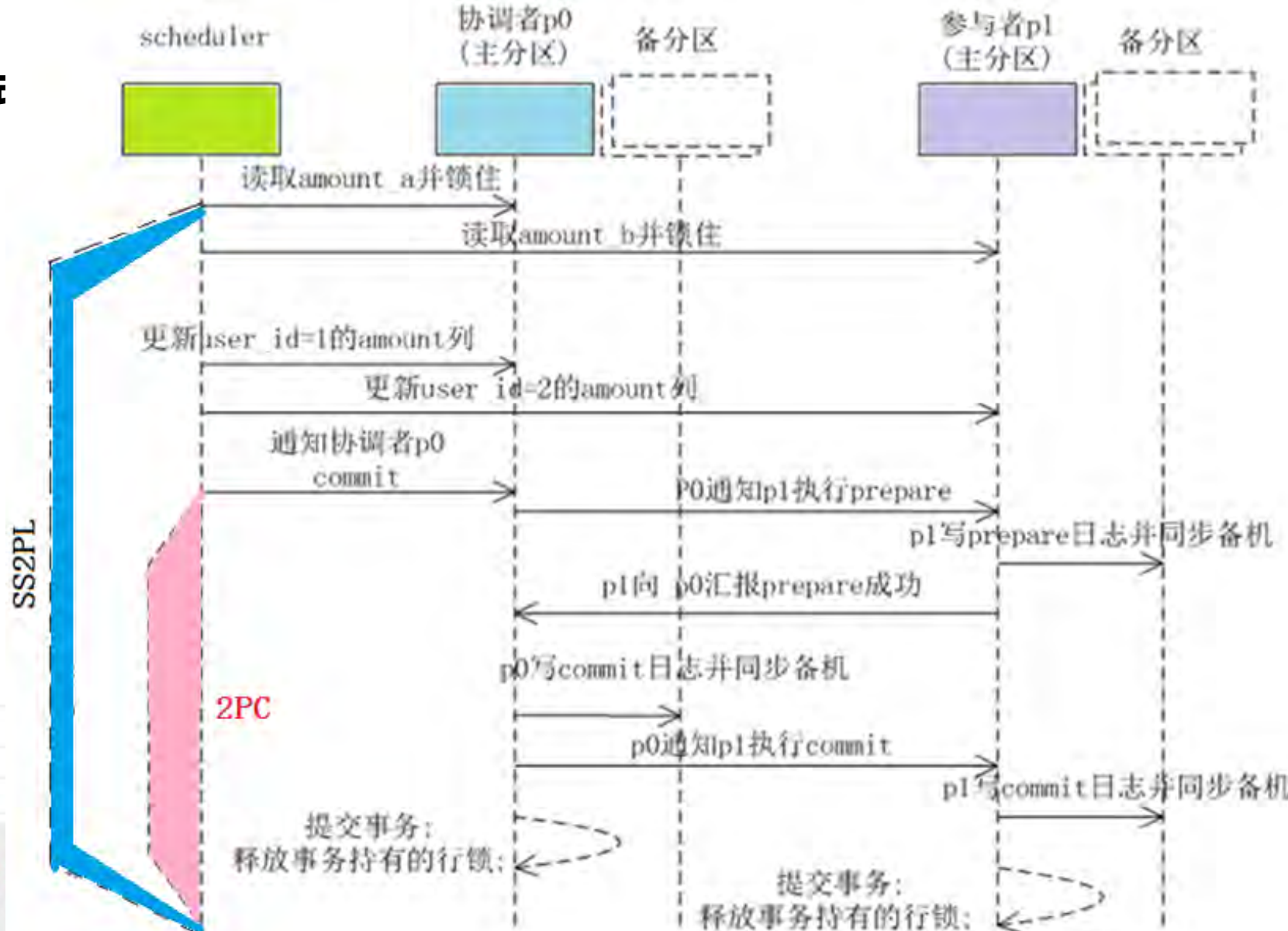


数据技术嘉年华
Data Technology Carnival



TDSQL分布式

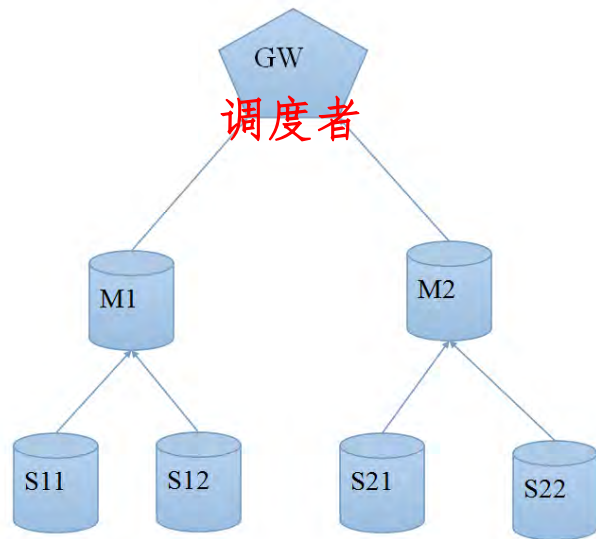
start transaction;
修改user1的金额;
修改user2的金额;
commit;



TDSQL分布式事务（金融云的需求和挑战）



- TDSQL的典型部署架构
 - 网关 (GW) -> **调度者**
 - 任意数量，通常每个DB实例配一个
 - group shard模式解析SQL语句
 - set (1主2备)
 - 1个 (noshard) 或多个 (group shard)
 - set1: {M1, S11, S12}; set2: {M2, S21, S22}
 - agent (每个DB实例1个)
 - 监控DB实例，完成集群下发的任务
- 网关支持用户发送多条写入SQL到多个set
 - 小表广播（一个基本静态的小表复制到所有set）
 - 多行插入语句
 - 多行更新删除
- 所有访问多个set的事务都是分布式事务
 - 内部自动识别，对用户透明
 - 两阶段提交



<http://www.cnblogs.com/qcloud1001/p/7472993.html>



第七届



数据技术嘉年华
Data Technology Carnival

TDSQL分布式方案...



TDSQL , 一直在努力...



第七届



数据技术嘉年华
Data Technology Carnival



目录

CONTENTS

TDSQL简介

TDSQL架构与分布式方案

TDSQL分布式事务处理

分布式事务处理技术



第七届



数据技术嘉年华

Data Technology Carnival



分布式数据库事务技术



	Xx DB	CockroachDB	Spanner	XxxxxBase
事务--ACID	支持	支持	支持	支持
并发控制	乐观/提交时检测冲突	乐观/MVCC	SS2PL/MVCC	SS2PL/MVCC
MVCC--多版本识别/ 全局唯一特性	事务ID	混合时间戳	物理时间戳 TrueTime	局部
MVCC-隔离特性	snapshot	write-snapshot	snapshot	snapshot
读写事务	乐观机制	乐观/MVCC	2PL	2PL
分布式事务提交/原子性	2PC	可避免2PC（事务状态记录）	2PC	2PC
外部一致性的读写	支持	支持	支持	存在不一致（因果序）
全局一致性的快照读	SI级别	SSI级别		SI级别
只读事务在备机/ follower上读	leader上读		支持	leader上读/ 备机弱一致性读
死锁	无死锁	无死锁	伤停等待（wound-wait）避免死锁	超时检测
预写日志/WAL	支持	支持	支持	支持
隔离级别	SI	SI/SSI	SI	RC

PostgreSQL, MySQL, Greenplum, Informix, etc

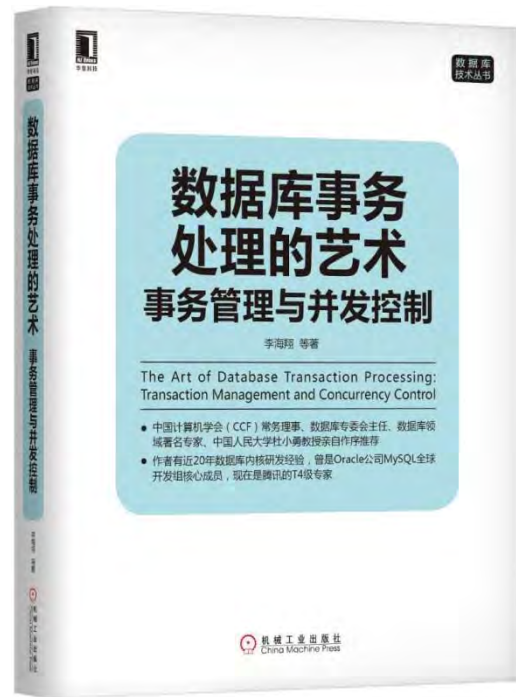
@那海蓝蓝 Blog: http://blog.163.com/li_hx/

2017 《数据库事务处理的艺术: 事务管理与并发控制》 2017年 机械工业出版社出版

2014 《数据库查询优化器的艺术: 原理解析与SQL性能优化》

本次分享的Ppt位于:

<http://pan.baidu.com/s/1jI6MBg6>



TDSQL



第七届



数据技术嘉年华
Data Technology Carnival

一个分享交流的地方

微信号: eyygle



Long Press QR Code To
Identify The Concern

长按二维码识别关注

扫一扫，加入我们，分享更多知识



第七届



数据技术嘉年华
Data Technology Carnival





THANKS

