



DTCC

2016中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2016

数据定义未来

SequeMedia
盛拓传媒

IT168.com

ChinaUnix

ITPUB

阿里海量数据迁移同步核心架构及最佳实践

付大超（千震）@阿里巴巴

2016.5.13



个人简介

- 付大超（花名：千震）
- 阿里巴巴数据库团队技术专家、DTS团队负责人
- 2012年加入阿里巴巴，目前负责DTS团队研发工作，曾负责阿里HBase的开发及维护工作，开发了阿里HBase集群高可用性系统，曾先后实习及工作于IBM、Cisco、淘宝。
- Email : qianzhen.fdc@alibaba-inc.com



2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

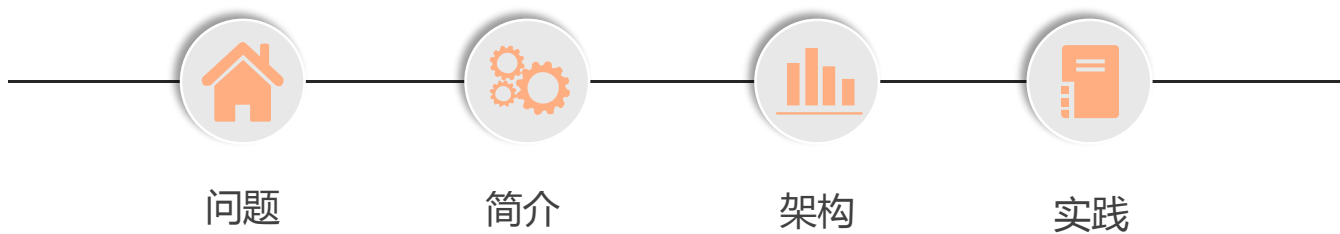
SequeMedia
数据传媒



ChinaUnix



目录 / CONTENTS





PART 1

问题



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

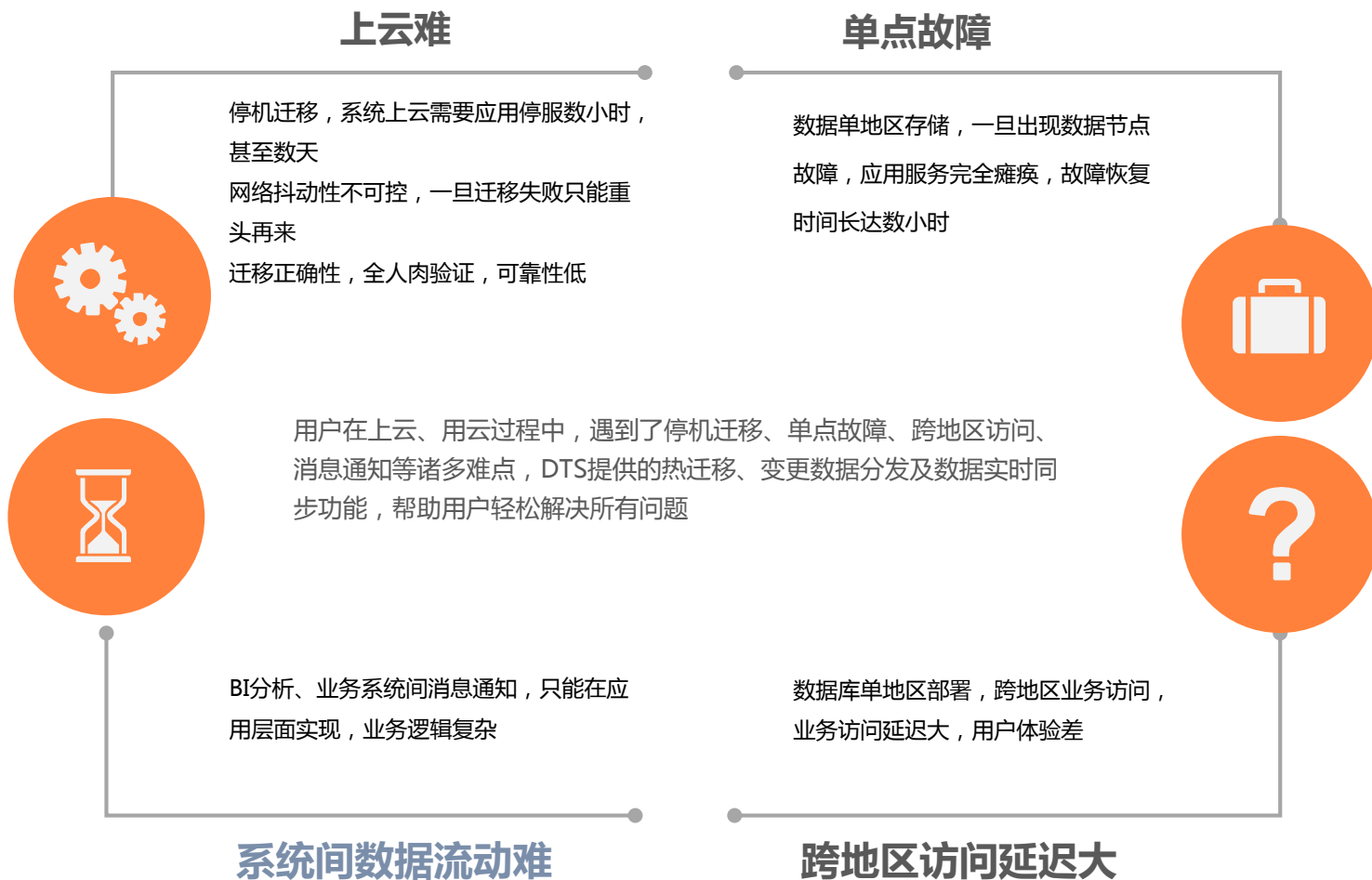
SequeMedia
数据传媒

IT168

ChinaUnix

IT-PUB

痛点



阿里典型业务挑战：

- 一键建站

要求一键完成PB级核心数据数天内从一个数据中心到1500公里外的另一个数据中心的迁移，并完成业务的无缝切换确保无任何数据丢失。

- 异地多活

阿里最新的数据架构中要求一个核心的数据中心全部挂掉不影响双11业务的正常运行。

事务强一致，低延迟，高可用

- 异构

要求支持多种数据存储产品

- 中美秒级同步

中美网络延时200ms, 要求秒级同步

- 增量订阅服务

要求MySQL、OceanBase、SQLServer、Oracle、HBase提供增量数据，供各类业务及大数据处理





PART 2

DTS简介



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
数据传媒

IT168

ChinaUnix

IT-PUB

DTS简介



DTS

数据传输服务(Data Transmission Service , 简称DTS) 支持以数据库为核心的结构化存储产品之间的数据传输，它是一种集数据迁移、数据订阅及实时同步于一体的数据传输服务。可以使用DTS轻松构建安全、可扩展、高可用的数据架构。

丰富多样

支持同异构数据源之间的迁移同步
支持全量增量多种传输方式，
增量迁移可以实现应用无缝切换

高性能

全量迁移性能高达70MB/s
实时传输性能高达30000TPS

安全可靠

任意一个服务节点发生故障，秒级恢复
安全传输协议，模块之间token认证
链路可靠性高达99.95%

简单易用

支持断点续传，任务中断可以断点续传
可视化管理界面，任务状态透明
链路即开即用，链路规格动态伸缩



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
数据传媒

IT68

ChinaUnix

ITPUB

DTS功能

- 数据迁移

支持多种数据存储产品，如Oracle、MySQL、SQLServer、PostgreSQL、PPAS、MongoDB、OceanBase等
支持异构数据源之间的结构对象及数据迁移，例如Oracle->MySQL、Oracle->PPAS
支持增量数据迁移，应用可以使用增量迁移实现应用无缝切换到目标RDS实例上
支持库表列三级映射、数据过滤等多种传输数据加工特性

- 实时同步

支持多种数据存储产品之间的实时同步
支持动态修改同步对象、修改同步速度等多种操作，充分满足业务需求

- 数据订阅

支持RDS多数据存储产品的增量数据实时订阅
支持在线查看增量数据、订阅时间回退、修改订阅对象等多种操作，充分满足业务需求

- 完善的监控体系

DTS展示同步速度、同步延迟等多种性能监控视图
可以对订阅状态、订阅消费延迟、同步延迟等设置阈值报警





PART 3

DTS架构



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

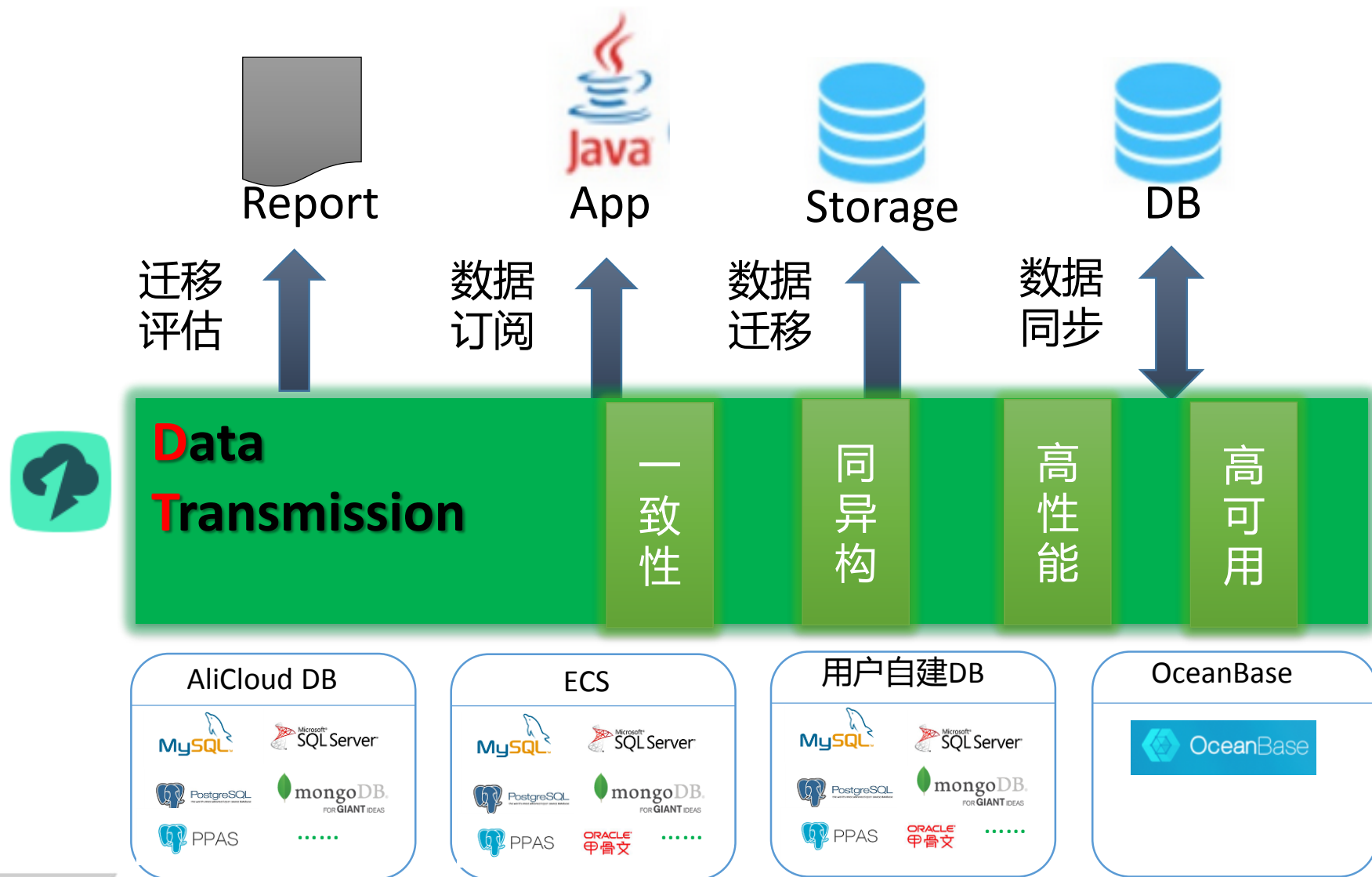
SequeMedia
数据传媒

IT168

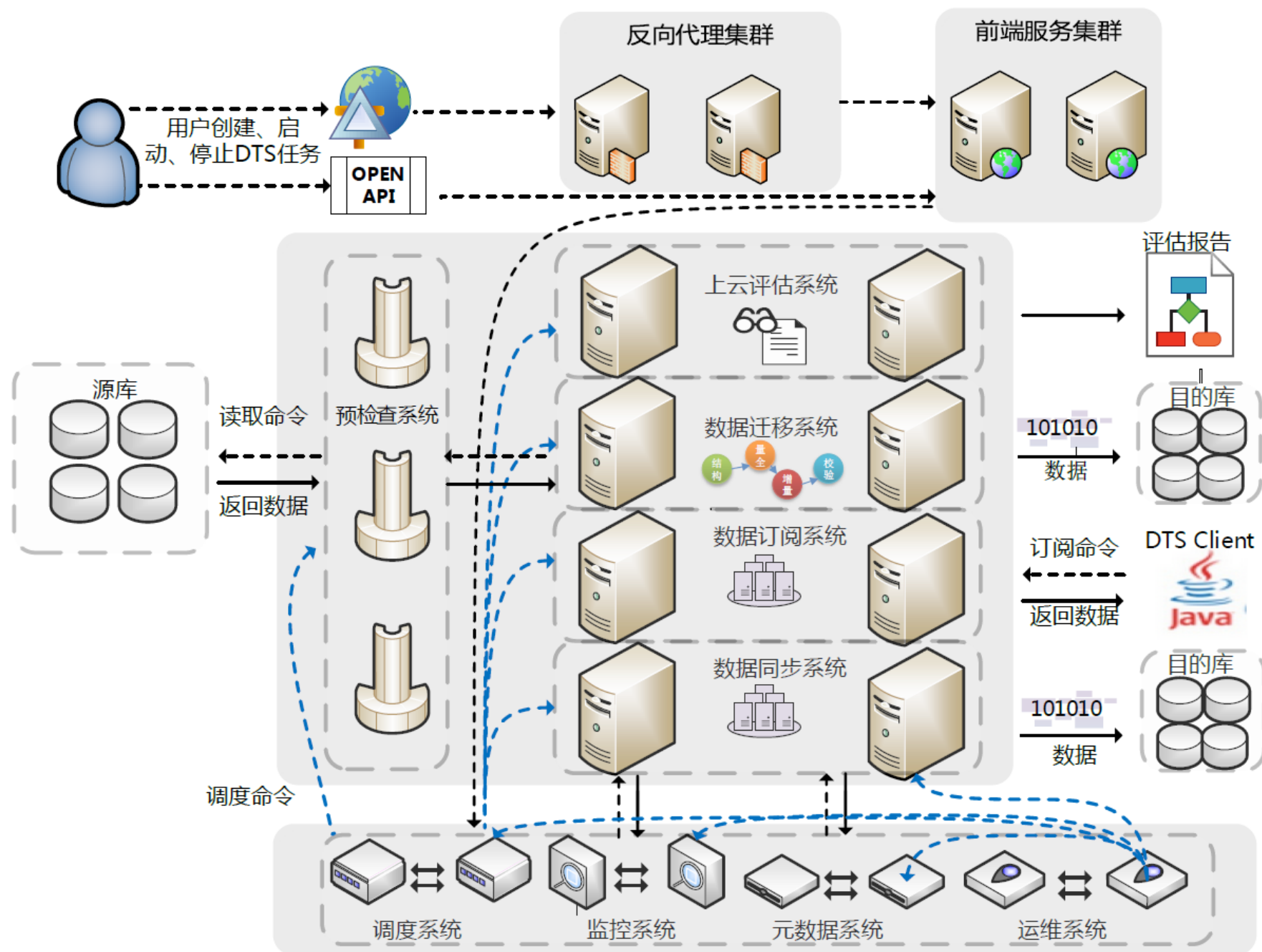
ChinaUnix

ITPUB

DTS介绍

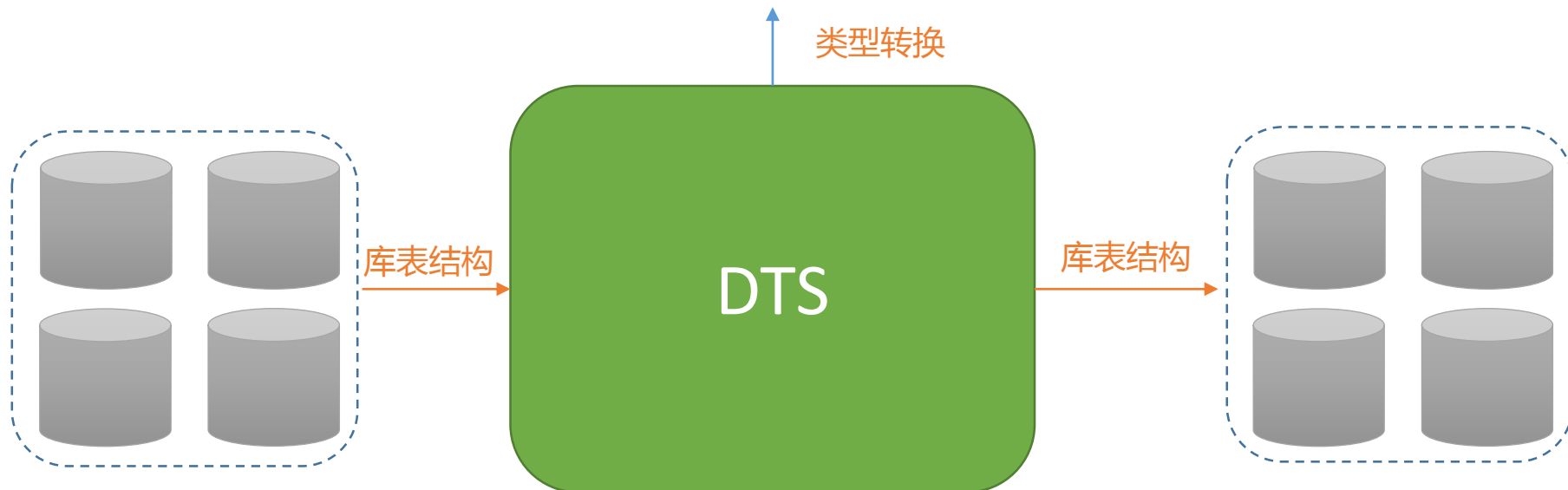


DTS架构

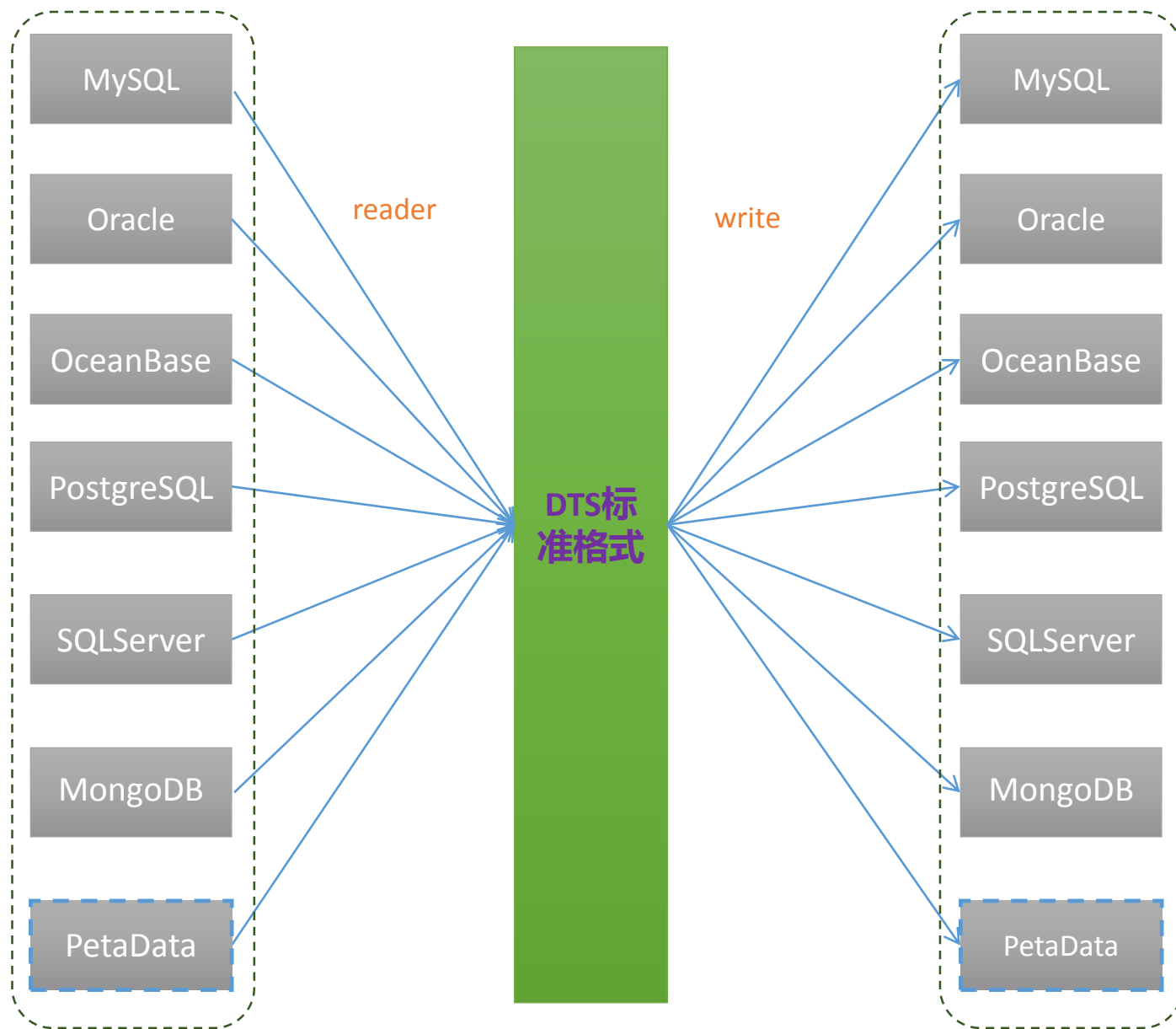


DTS结构迁移

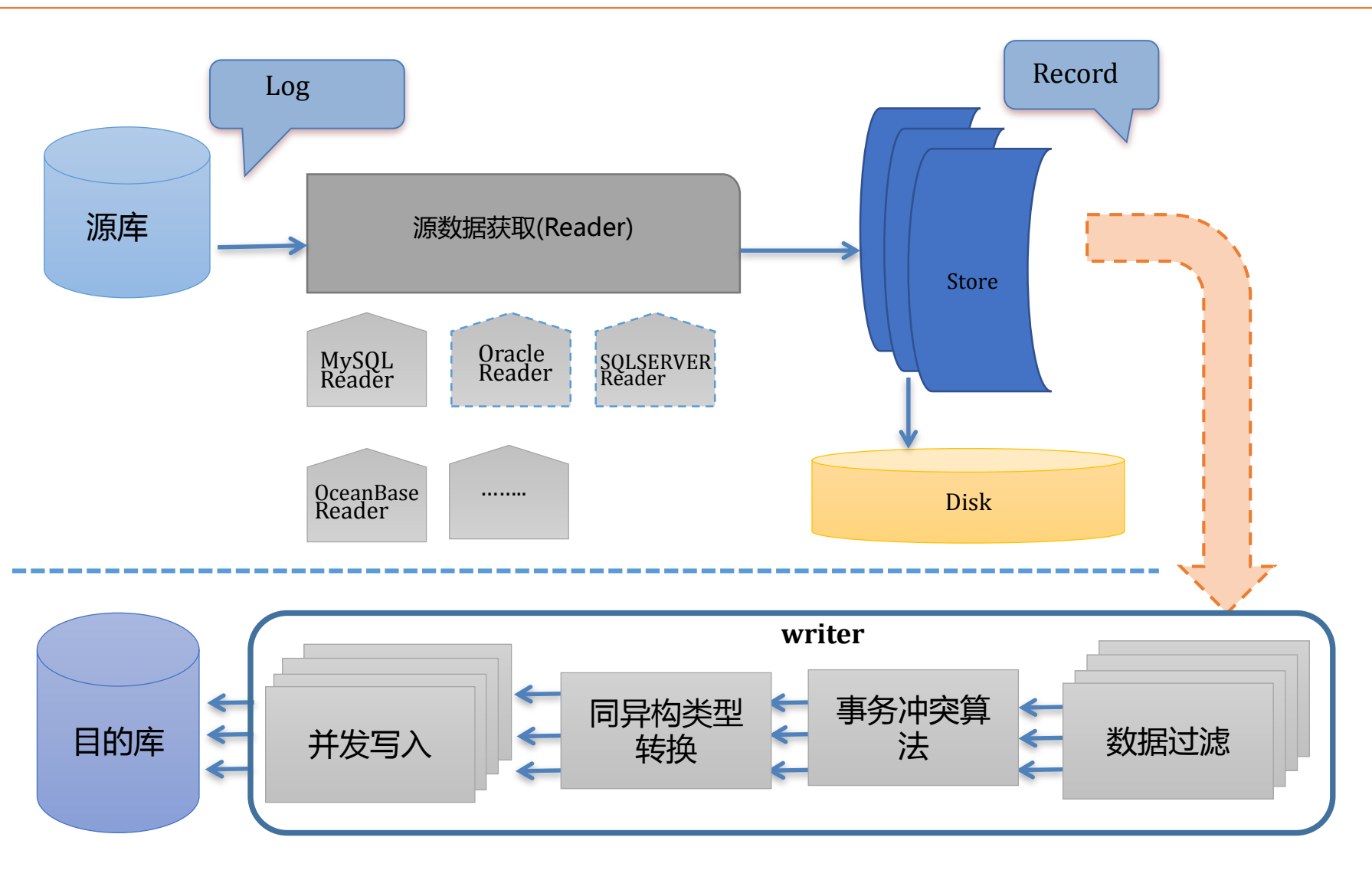
数据源	类型	类型	类型
Oracle	VARCHAR2	NVARCHAR2	LONG
Mysql	VARCHAR	VARCHAR	LONGTEXT



DTS全量迁移



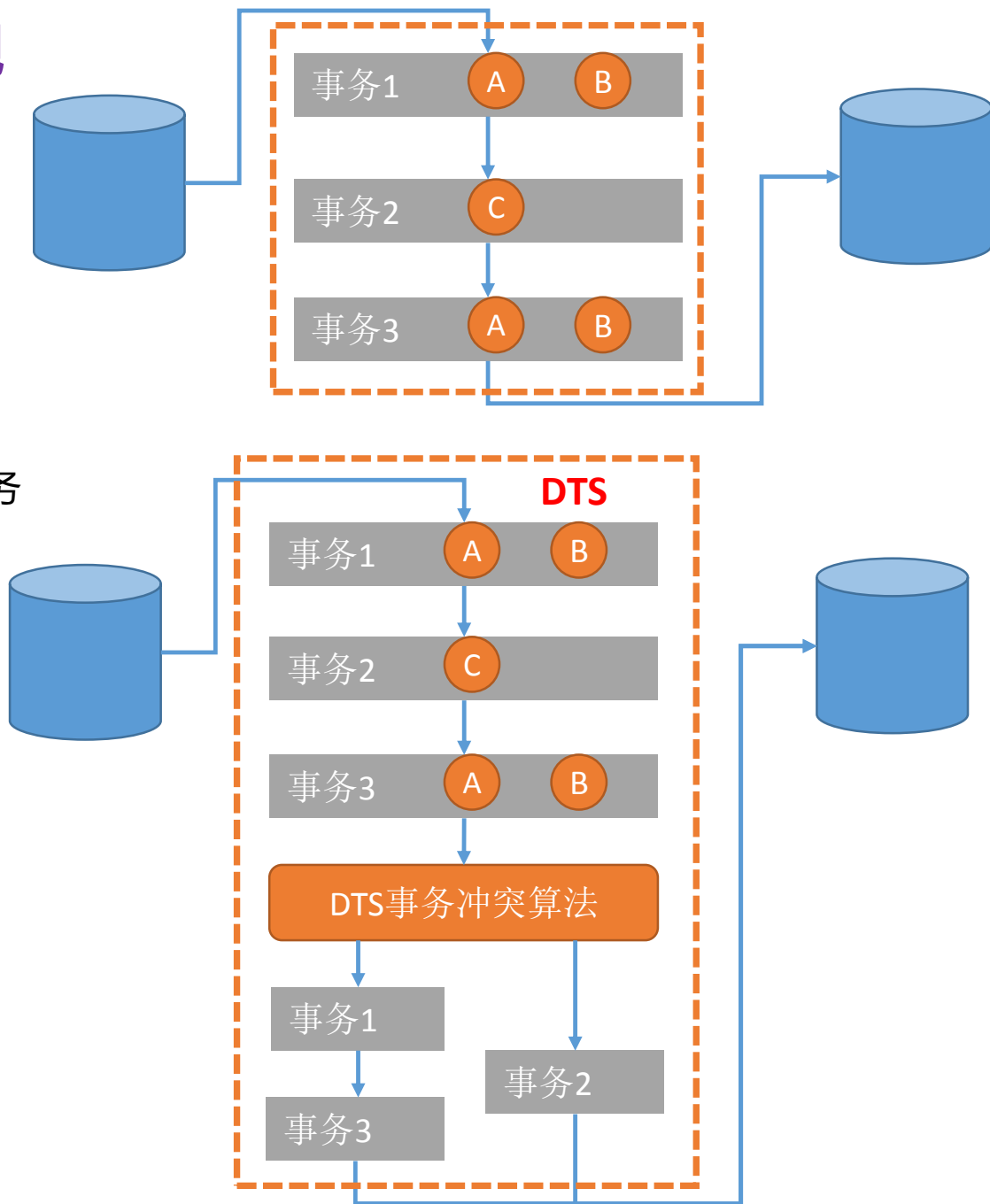
DTS增量&同步技术实现



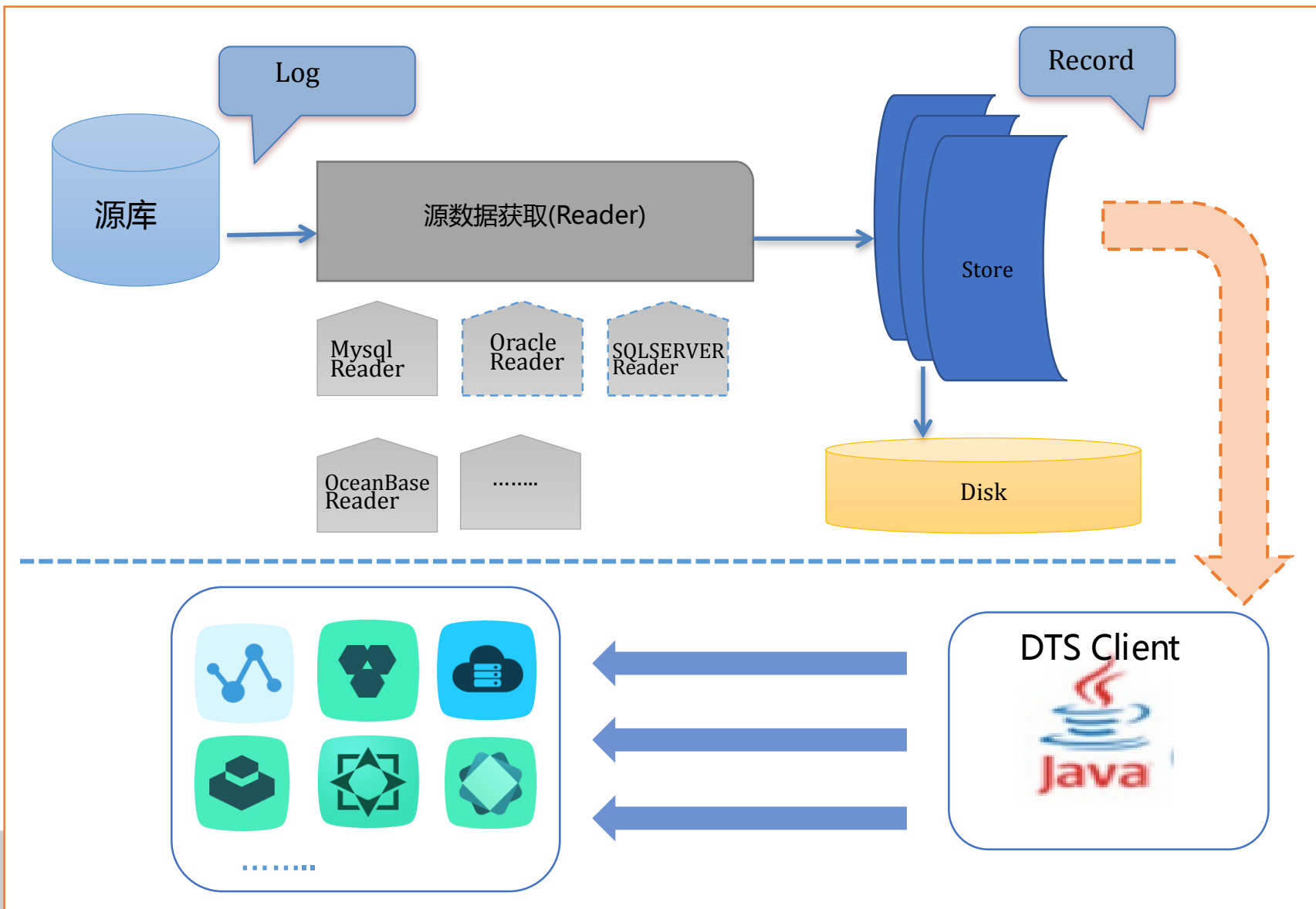
DTS增量&同步技术实现

•增量&同步Writer

- 高性能
 - 低冲突场景10万rps
- 支持多种数据库
 - JDBC模式：RDBMS
 - K/V模式：NoSQL、MQ
- 并发问题
 - 性能需要：串行事务->并发事务
 - 实时检测事务之间是否冲突
- 防数据回流
 - 多向同步时，避免重复写入
- 支持DDL同步



DTS订阅技术实现



DTS可用性

- 升级

目前portal、API升级对于用户无感知，因为Portal和API各有两台机器。

目前Master、Node升级对于已经运行的任务不影响，可能会影响即将启动的任务，但时间非常短，可以视作对于用户无感知。

目前预检查、结构迁移、全量迁移、增量迁移升级对于已经运行的任务不影响，也不会影响即将运行的任务，时间非常短，可以视作对于用户无感知。

- 任务级别

所有模块(预检查、结构迁移、全量迁移、增量迁移、Store、Congo)的单个任务会自动容灾

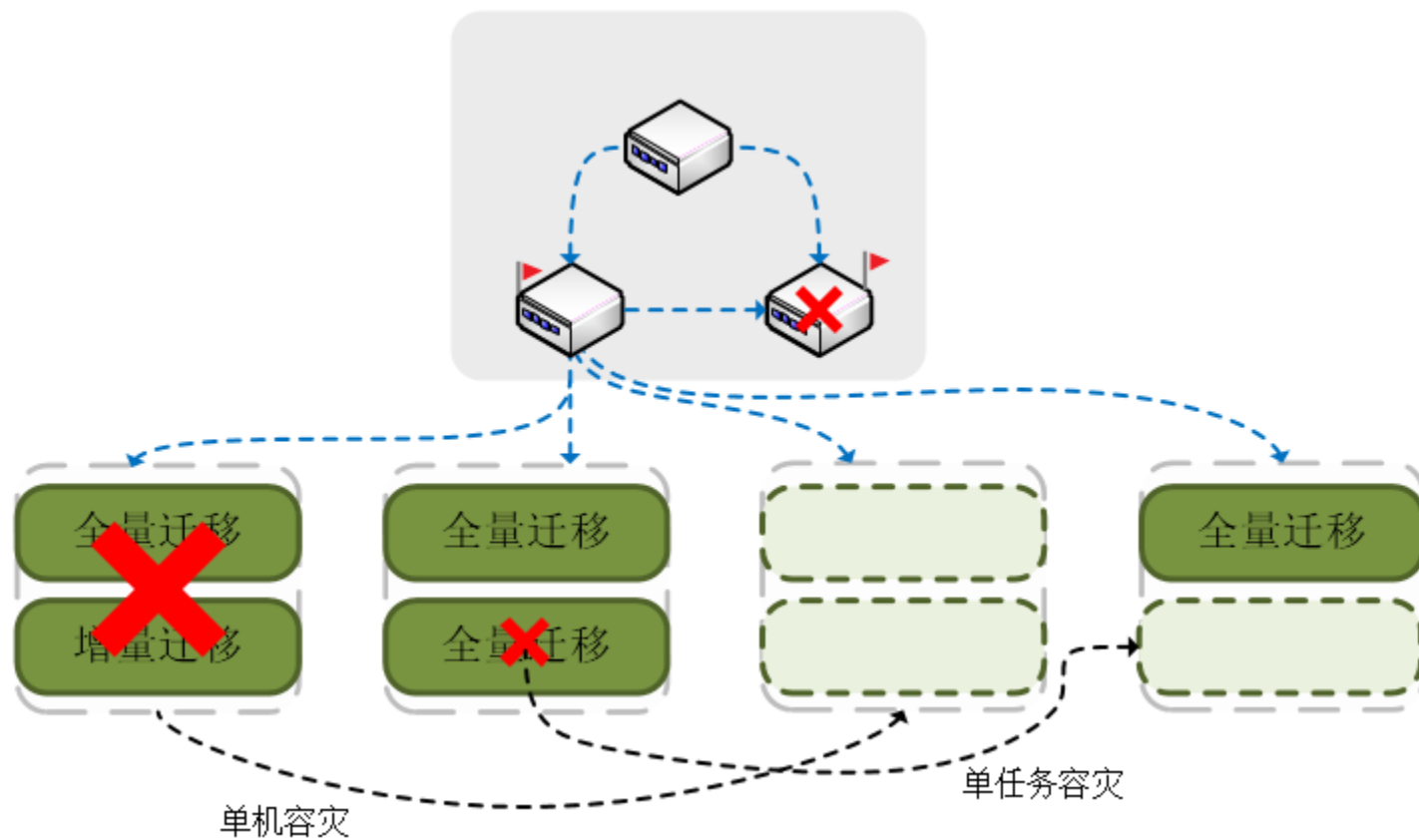
- 主机级别

主机宕机后，上面的任务会自动容灾。

- 监控

对失败的任务进行报警







PART 4

阿里实践



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
数据传媒

IT168

ChinaUnix

IT-PUB

DTS使用概况

公有云：为几百W用户提供上云迁移及阿里云上各个地区的迁移、异地实时同步、数据订阅功能。

专有云：提供输出给众多核心项目支持迁移及异地实时同步

阿里、蚂蚁：支持超过1W+DTS任务，超过PB+数据迁移及同步。



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
数据传媒

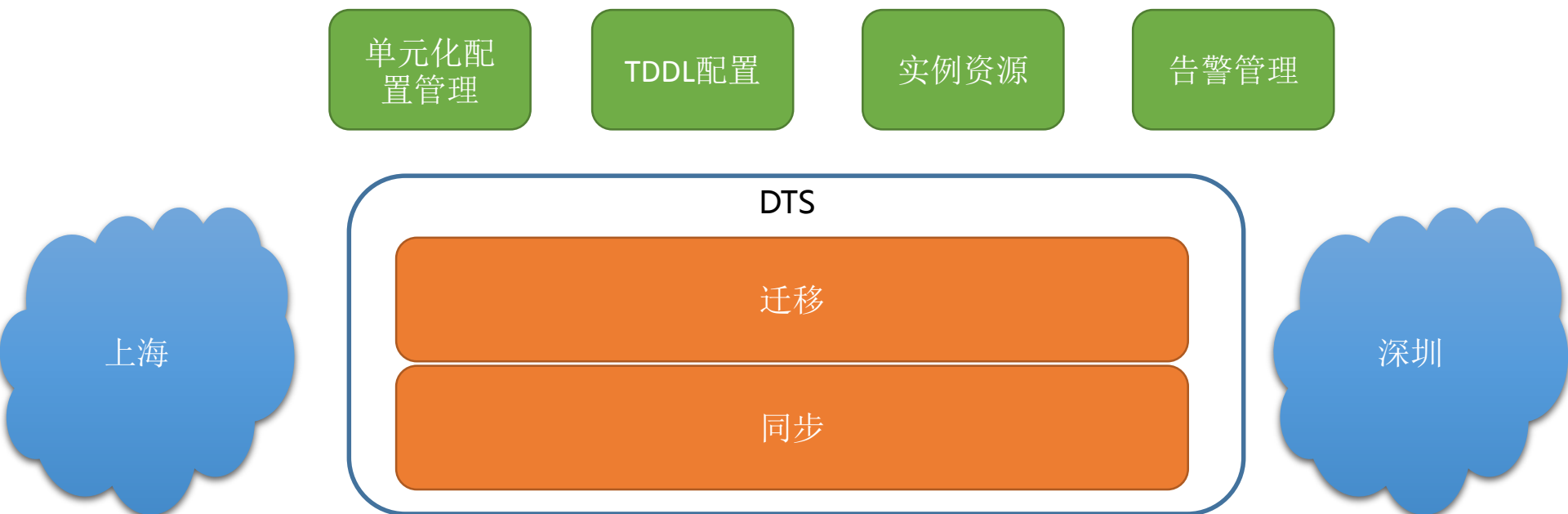
IT68

ChinaUnix

IT-PUB

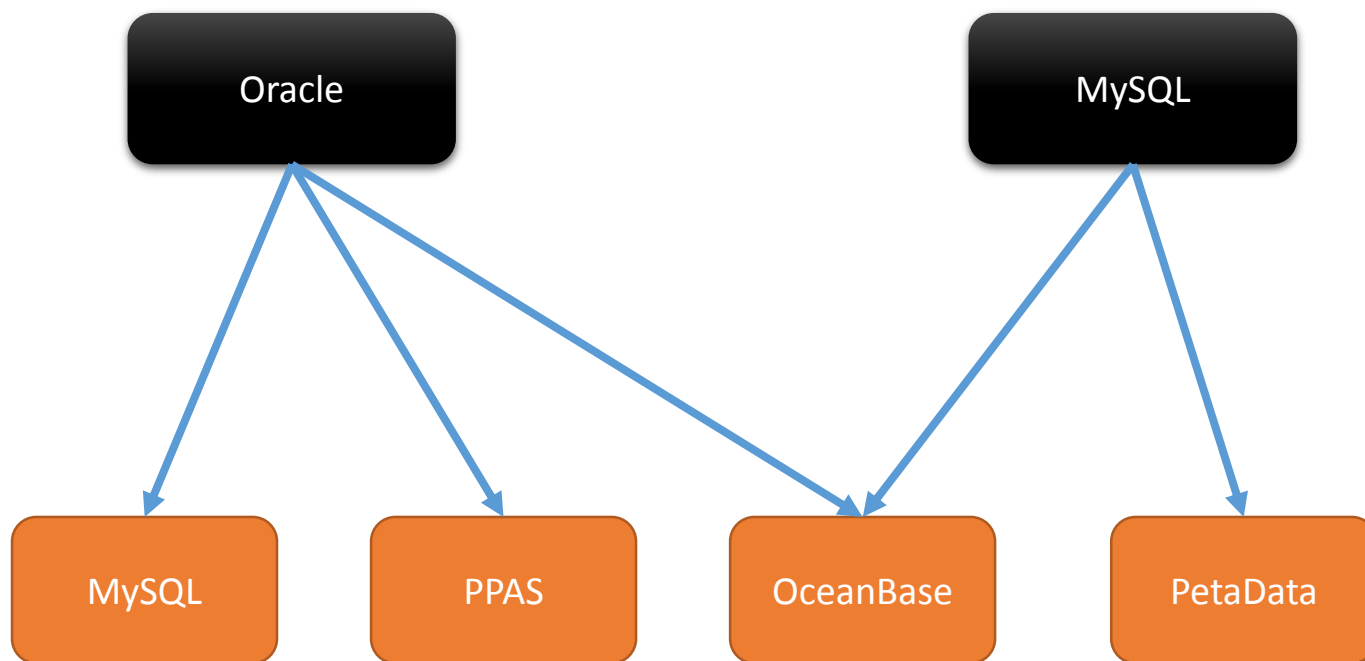
阿里核心场景实践：一键建站

双11上百个应用、上千个实例、PB级数据、1000公里+、数据一致

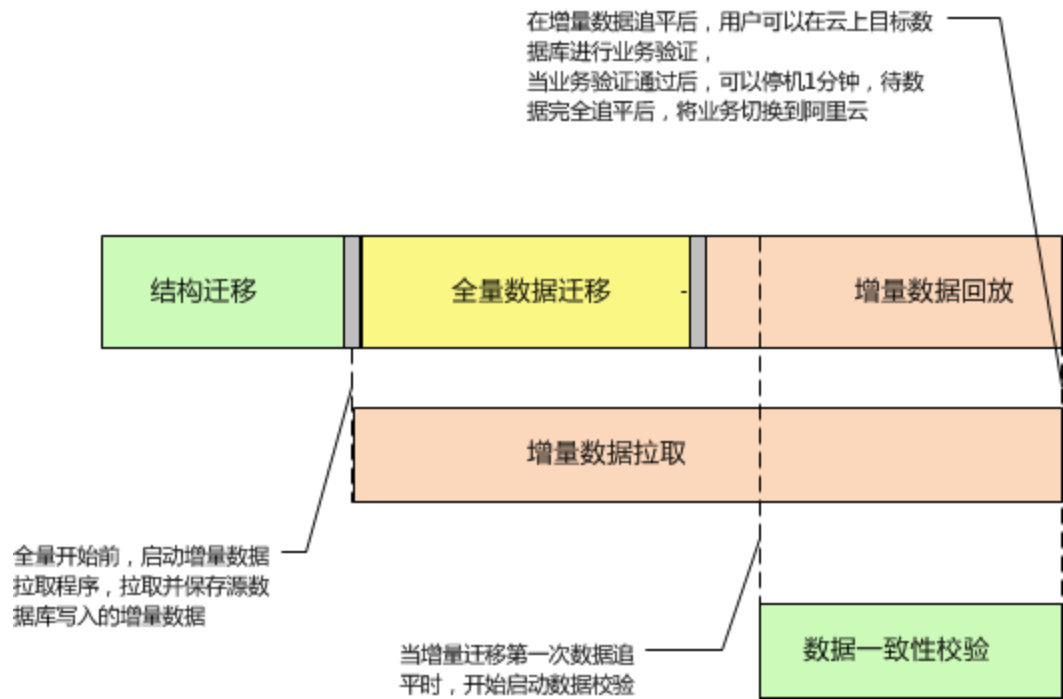


阿里核心场景实践：异构迁移

通过DTS的结构迁移、全量迁移、增量迁移、全量校验、增量实时校验完成交易、会员.....等迁移



阿里云核心场景实践：上云迁移



热迁移

DTS支持增量迁移，上云迁移过程中，本地业务可以继续提供服务，从而最大程度降低数据迁移期间的应用停机时间
业务在云上验证正确后，可以选择停机数分钟，待两边数据完全追平后切换到阿里云

断点续传

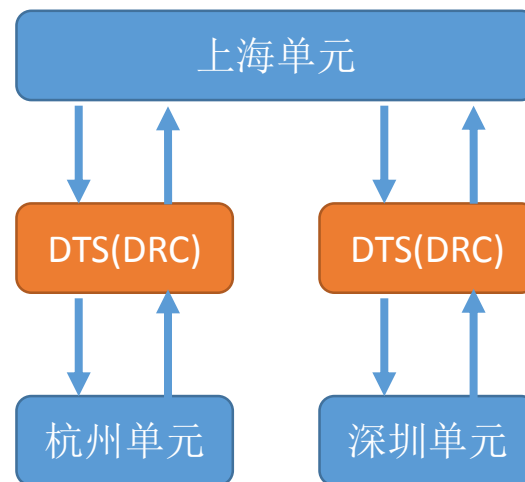
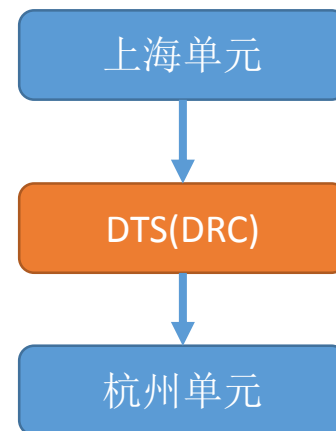
数据传输服务支持任务断点续传，有效提高恶劣网络环境下迁移成功率，避免因网络等因素导致任务中断后，数据需要重头迁移

数据一致性校验

DTS支持迁移数据一致性校验，包括全量校验及增量数据实时校验，快速、准确地验证迁移数据的一致性

阿里核心场景实践：异地多活

- 淘宝核心交易库
 - 双十一数据量
 - 全天同步总量：百TB级别（全网6倍）
 - 高峰增量流量：Gbps级别
 - 增量消费的下游
 - 实时媒体大屏
 - 实时商业分析
 - 实时搜索
 - 实时备份
 - 离线分析平台
 - 含多个子任务
 - 2个异地实时备份链路（异地多活）
 - 上海到杭州
 - 上海到深圳



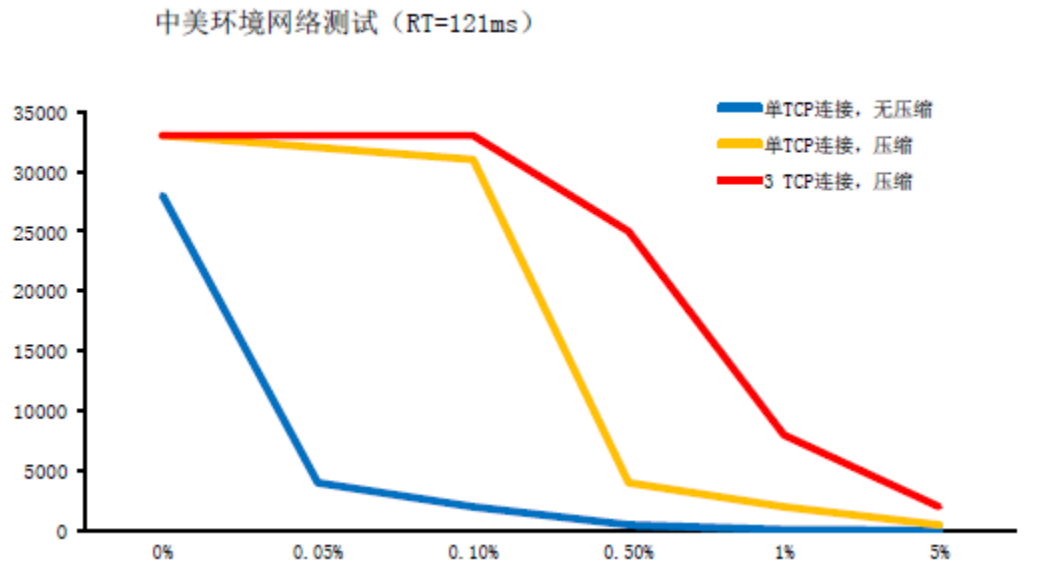
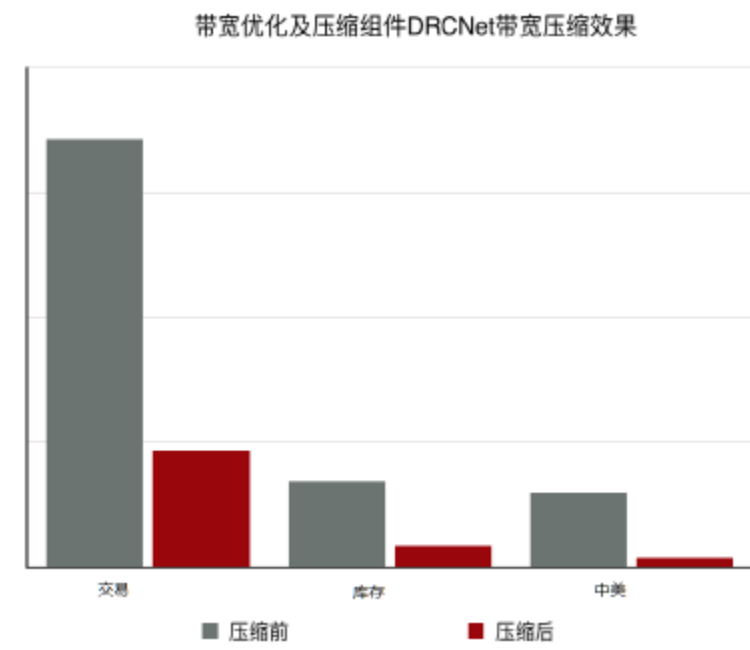
阿里核心场景实践：中美同步

Net组件

- 实时压缩，解决骨干网、主机网卡资源瓶颈问题
- 压缩后延时增加不高于5ms，CPU增加不高于单核100%(占满2G带宽时)
- 多TCP协同传输，降低丢包对TCP传输带宽的影响
- SSL加密

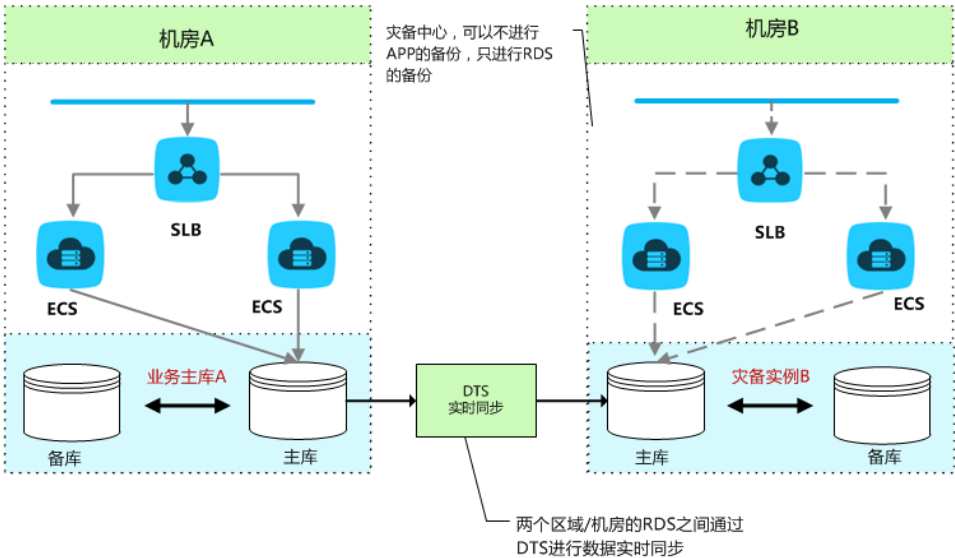


数据大小		LZ4		snappy		QuickLZ		miniz	
		压缩时间us	解压时间us	压缩时间us	解压时间us	压缩时间us	解压时间us	压缩时间us	解压时间us
1K	压缩时间us	20	14	52	134	52	134	52	134
	解压时间us	4	4	49	52	49	52	49	52
	压缩后字节数(Byte)	691	715	753	526	753	526	753	526
10K	压缩	10	28	0:00:00	201	0:00:00	201	0:00:00	201
	解压	4	4	59	101	59	101	59	101
	压缩后字节数	1495	1863	1681	1158	1681	1158	1681	1158
1M	压缩	798	1314	6960	11474	6960	11474	6960	11474
	解压	668	318	5666	6355	5666	6355	5666	6355
	压缩后字节数	125080	178042	133579	91476	133579	91476	133579	91476
16M	压缩	14992	23837	114562	194631	114562	194631	114562	194631
	解压	10949	5553	99758	106428	99758	106428	99758	106428
	压缩后字节数	2194289	3033098	2315524	1623278	2315524	1623278	2315524	1623278
压缩(压缩时间us/压缩后字节数)		19/1029	14/1029	46/1033	139/642	46/1033	139/642	46/1033	139/642
解压(解压时间us)		3	1	8	50	8	50	8	50

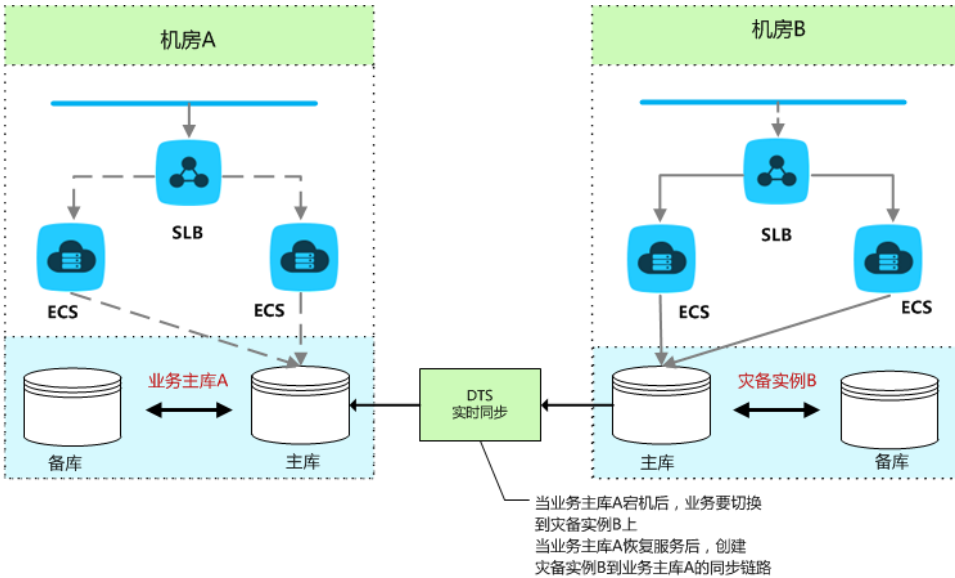


阿里云核心场景实践：异地灾备

切换前



切换后



支持地区

支持杭州、北京、上海、深圳、青岛、香港、美国硅谷及新加坡等8个站点
支持以上8个站点的RDS实例之间的数据实时同步

故障恢复

当源跟目标RDS网络延迟在30ms以内，
源/目标RDS没有带宽限制，
目的RDS没有写入性能瓶颈，
源RDS实例的TPS（每个TPS只有单条记录更新）
在10000以内时，同步延迟可达1秒以内

当业务主库故障时，业务恢复的RPO 0秒，RTO秒级



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

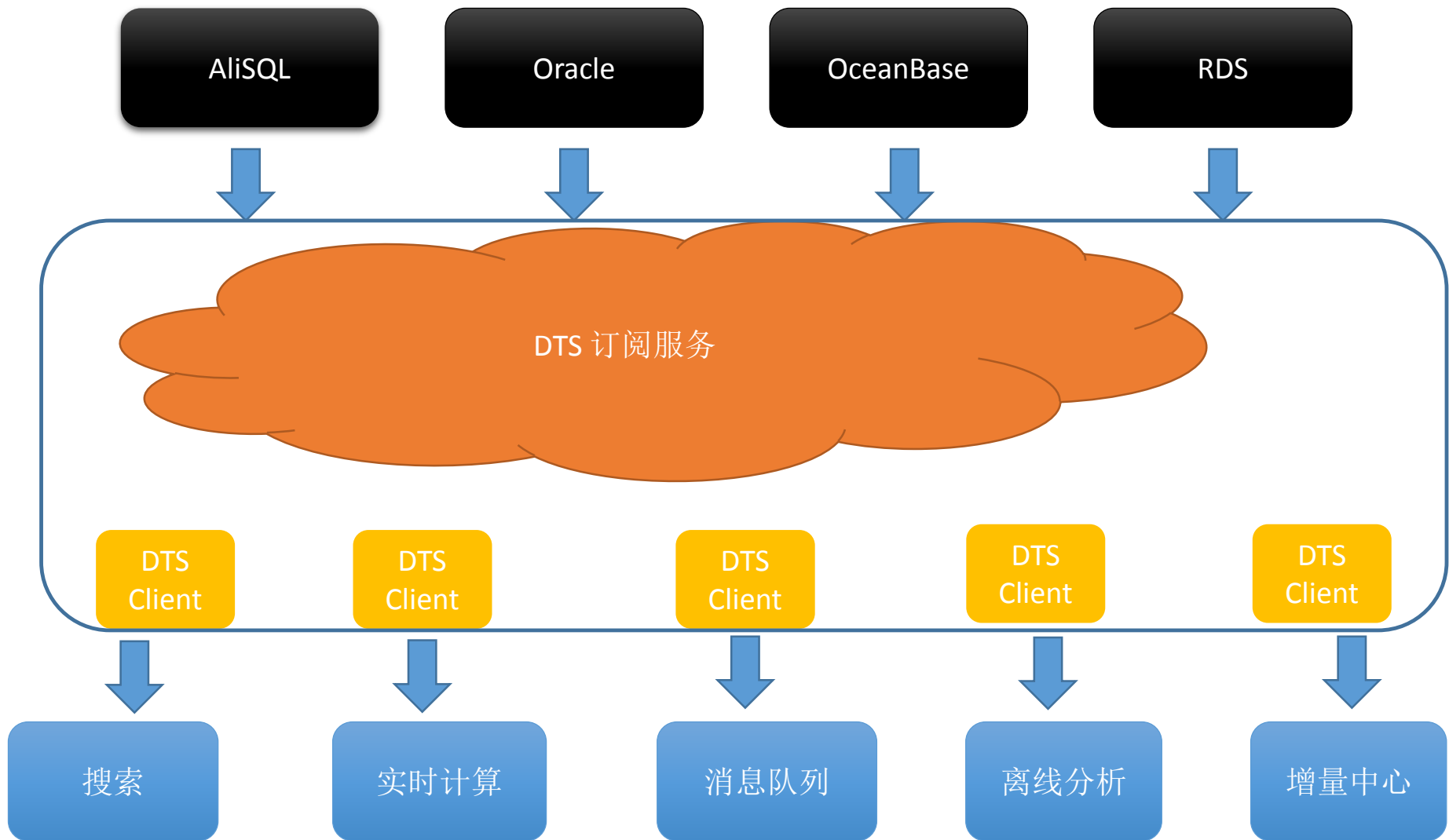
SequeMedia

IT68

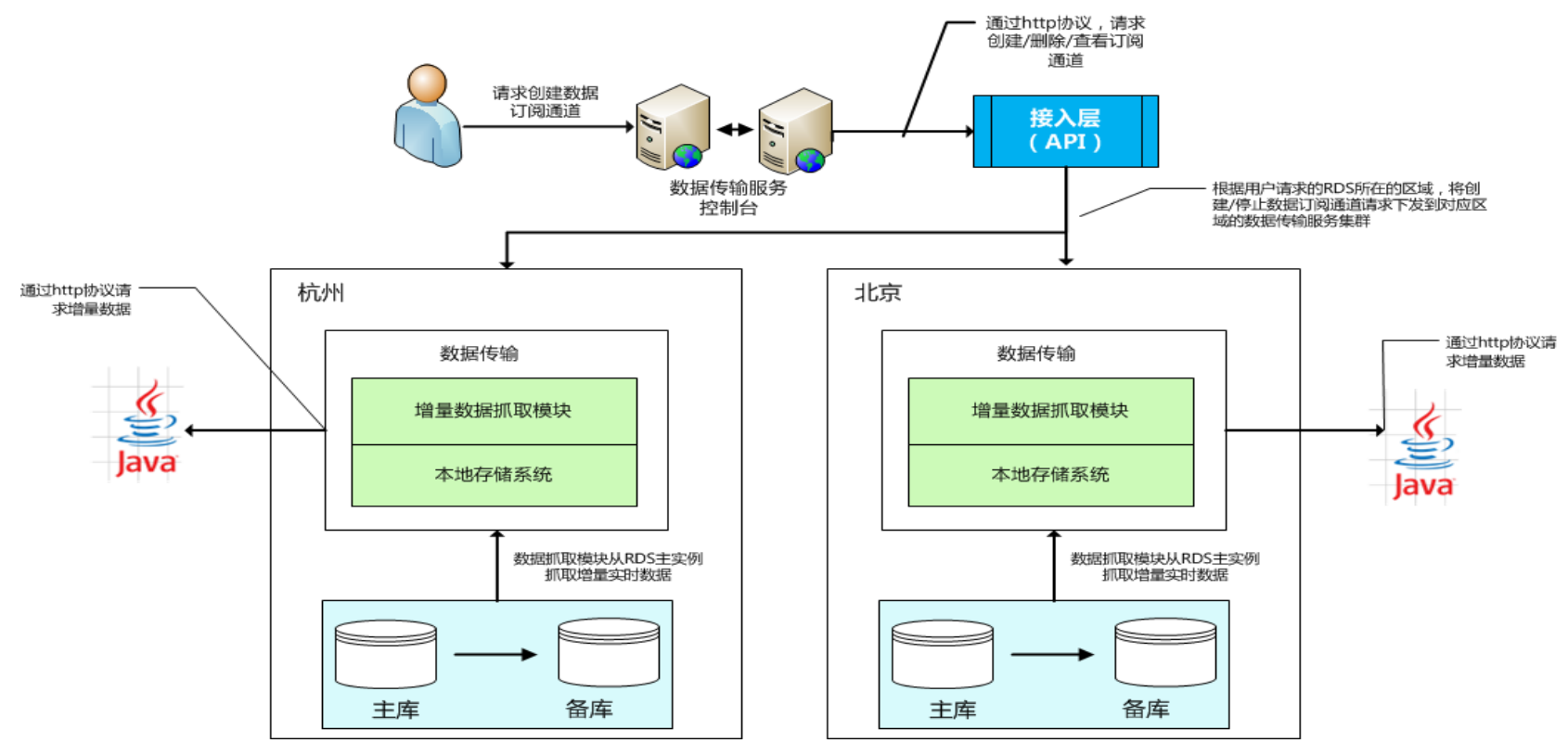
ChinaUnix

ITPUB

阿里核心场景实践：增量订阅服务



阿里云核心场景实践：数据订阅



本地灾备

通过订阅云上增量变更数据，封装成SQL语句，并回放到本地DB，在本地IDC构建阿里云业务灾备中心，保证业务高可用

BI分析

订阅云上增量变更数据，并按照业务分析逻辑进行处理后导入到数据仓库中，在不影响业务主库的情况下，满足BI分析需求

消息通知

业务系统之间的消息通知机制，可以通过订阅业务A的变更数据，触发业务B中相关业务操作。简化业务耦合逻辑，避免实现复杂的通知逻辑

阿里云DTS VS AWS DMS

	阿里云DTS	AWS DMS
产品定位	专注于以数据库为核心的结构化存储产品之间的数据传输，打造专业的数据流产品	数据库之间的数据迁移 实现数据库热迁移，最大程度地为依赖迁移数据库的应用程序减少停机时间
功能列表	数据迁移 实时同步 数据分发 数据一致性校验 上云评估	数据迁移 实时同步
支持数据源	MySQL (增量)、SQLServer (增量)、Oracle (增量)、PostgreSQL (增量)、PPAS (增量)、MongoDB、OceanBase	MySQL (增量)、SQLServer、Oracle (增量)、PostgreSQL、Aurora
ETL	支持简单ETL： 迁移对象选择 库、表、列名映射 数据过滤	支持简单ETL： 库、表名映射



Q&A

谢谢！



DTCC

2016年中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2015

SequeMedia
数据传媒

IT168

ChinaUnix

IT-PUB



THANKS

SequeMedia
盛拓传媒

IT168.com

ChinaUnix

ITPUB