

让大数据流动起来

数据蜂巢架构演进之路

贺思远

CONTENTS

Introduction

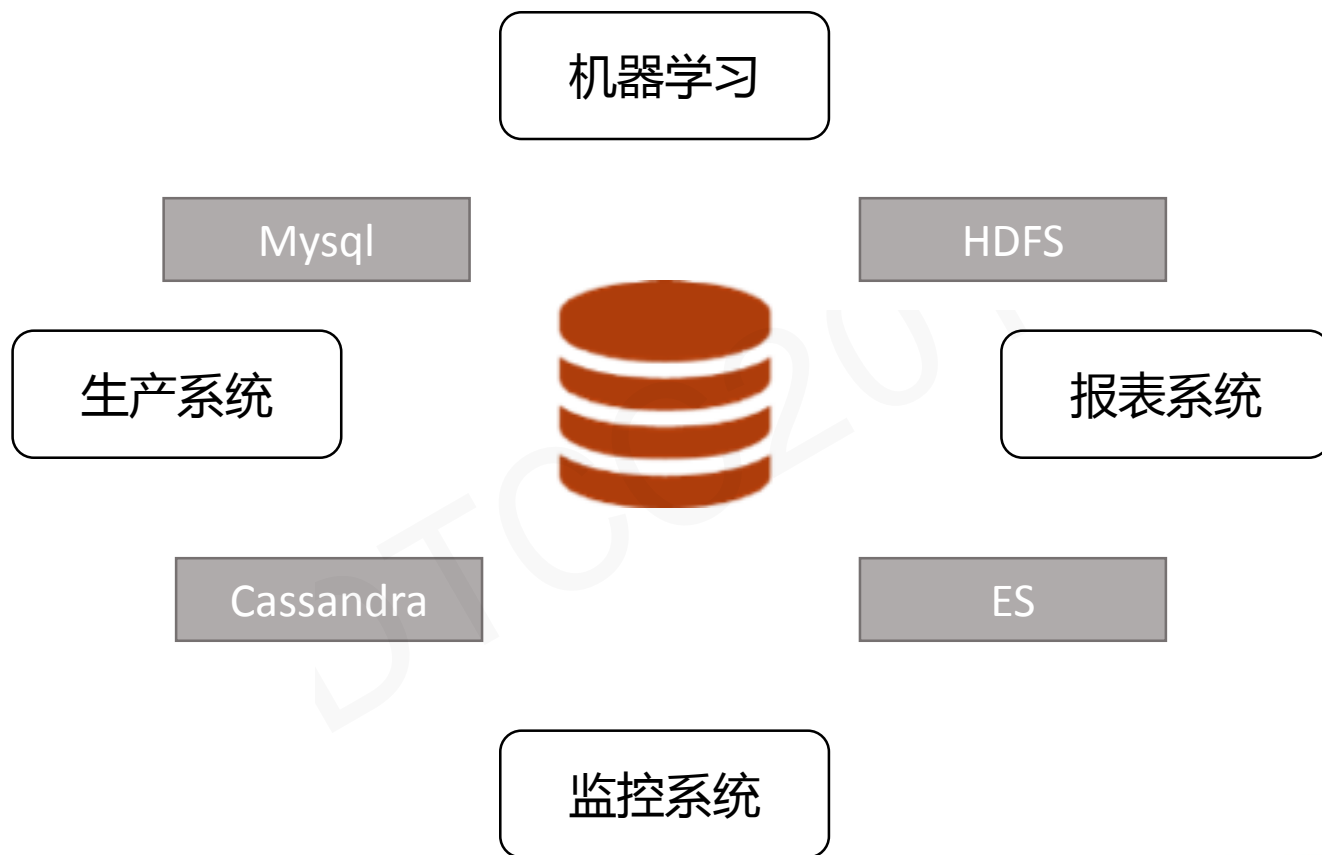
□ 背景

□ 演进

□ 架构

□ 现状

背景



CONTENTS

Introduction

□ 背景

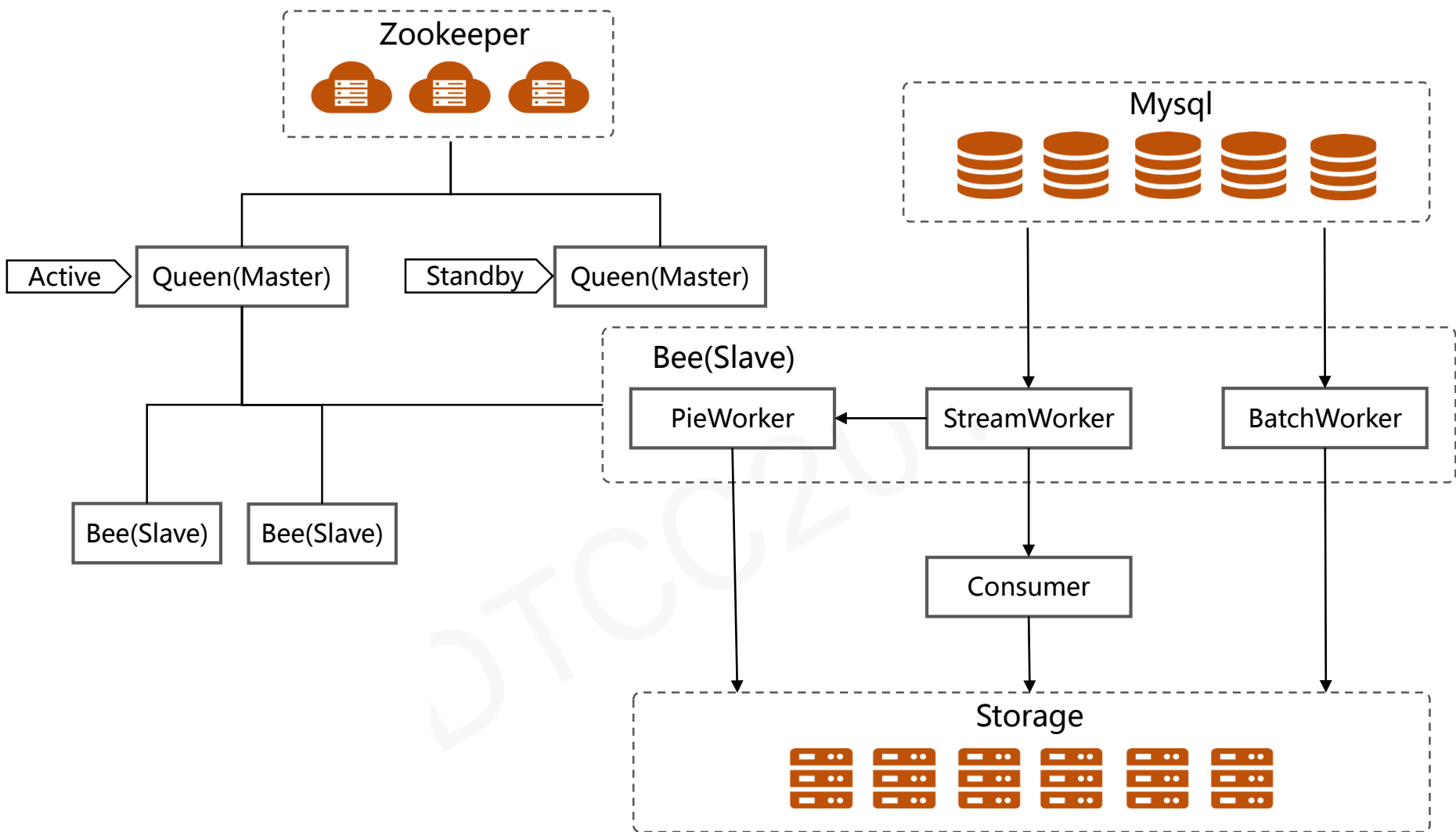
□ 演进

□ 架构

□ 现状

架构





架构-作业

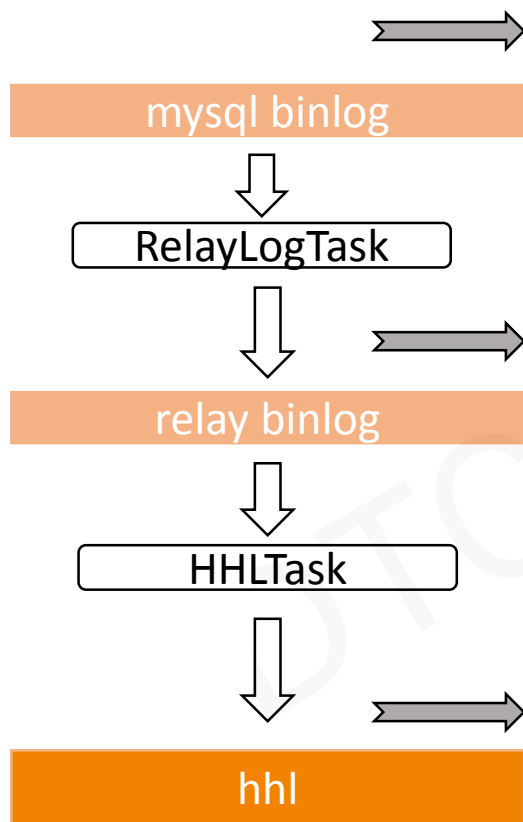
- Job
 - 批量同步-BatchJob
 - 数据实时订阅-StreamJob
 - 实时同步-PieJob
- 实时&离线
- 分片

架构-BatchJob



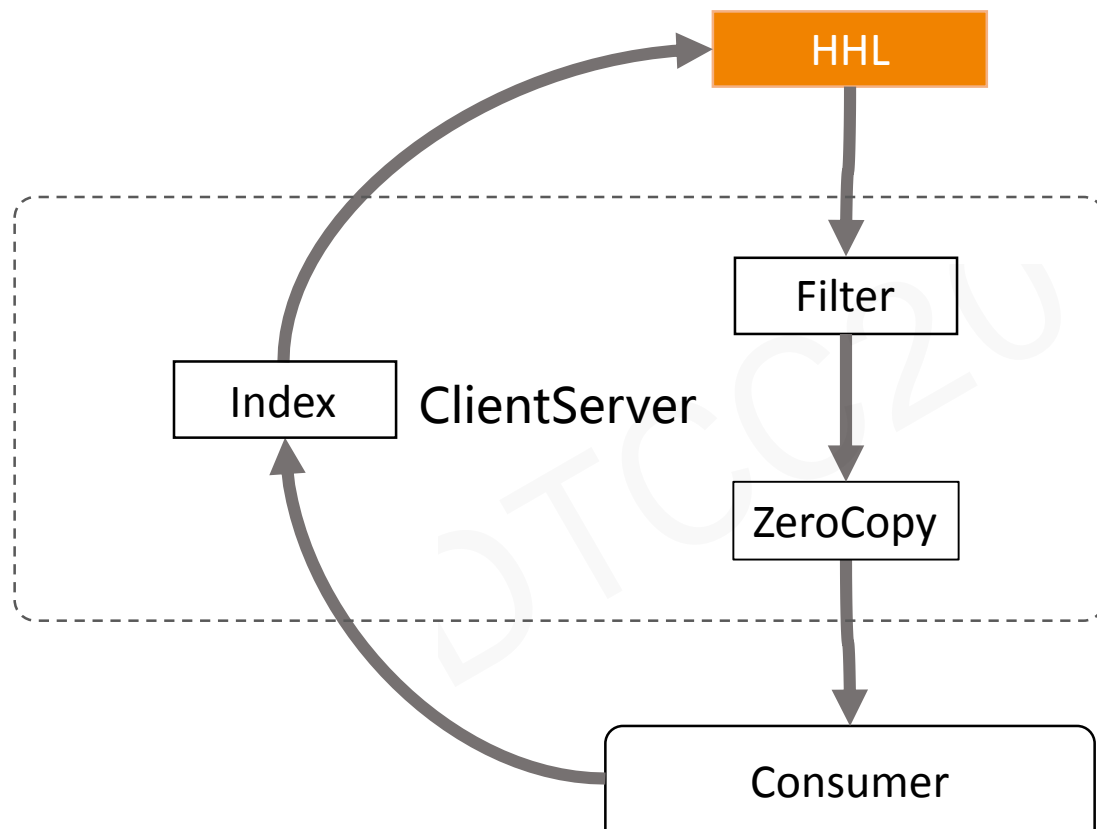
- Fetcher抽取数据，Storage缓存，Sinker用于数据处理

架构-StreamJob



- RelayLogTask
- HHLTask
- HHL

架构-Client



架构-Client

- Pull
- serverId+binlogPosition+time
- 支持多种并行模式
 1. 串行，完全支持事务
 2. 事务级并发，保证事务原子性，只有单表操作的事务进行并发，多表操作的事务继续串行。
 3. 表级并发，保障同一表操作的有序
 4. 行级并发，保障同一主键的操作有序

架构-集群

高可用

数据本地
性

负载均衡

CONTENTS

Introduction

□ 背景

□ 演进

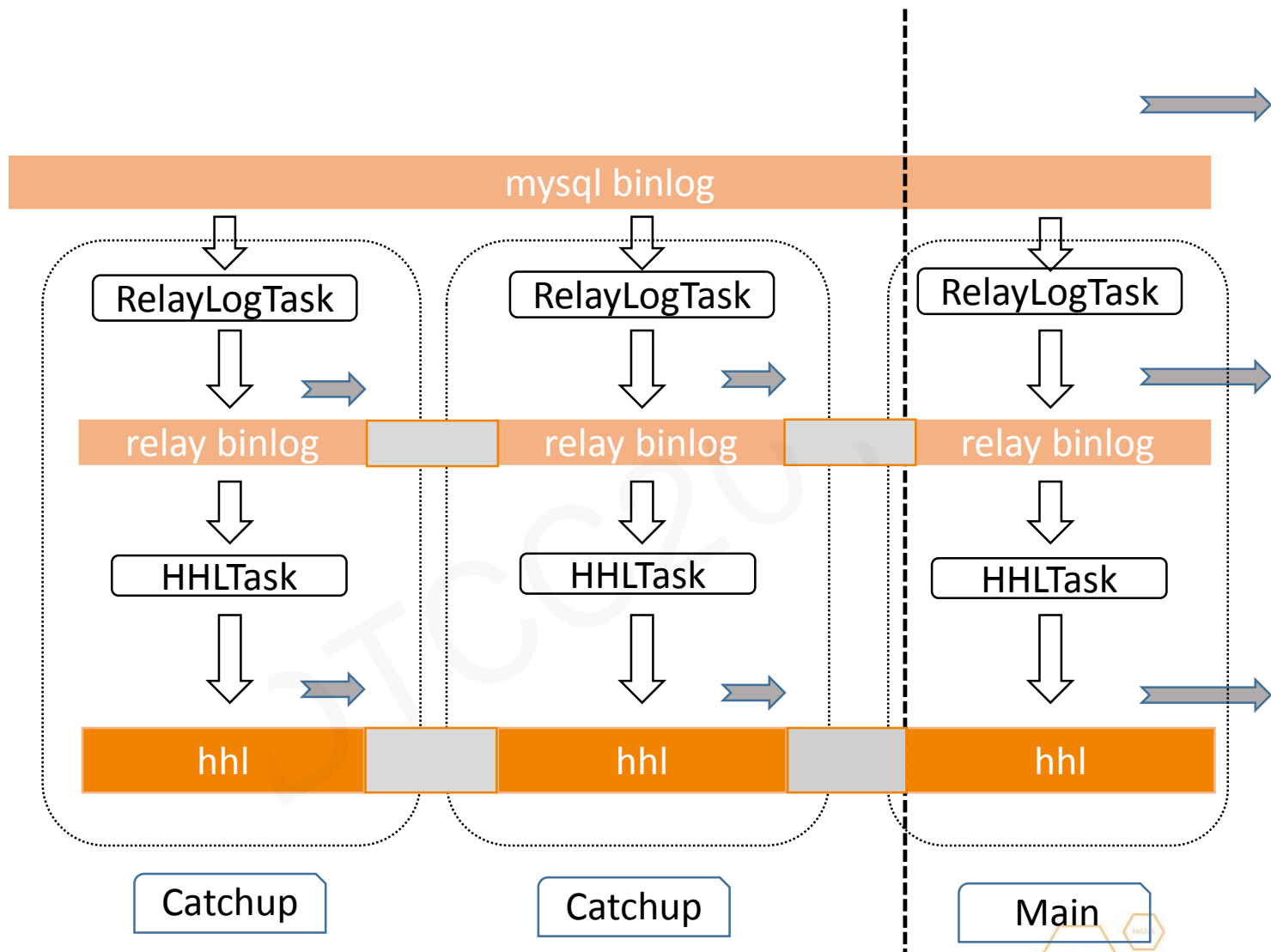
□ 架构

□ 现状

演进-HHL文件丢失

- 复本
- 数据补全

DTCC2018



演进-元数据

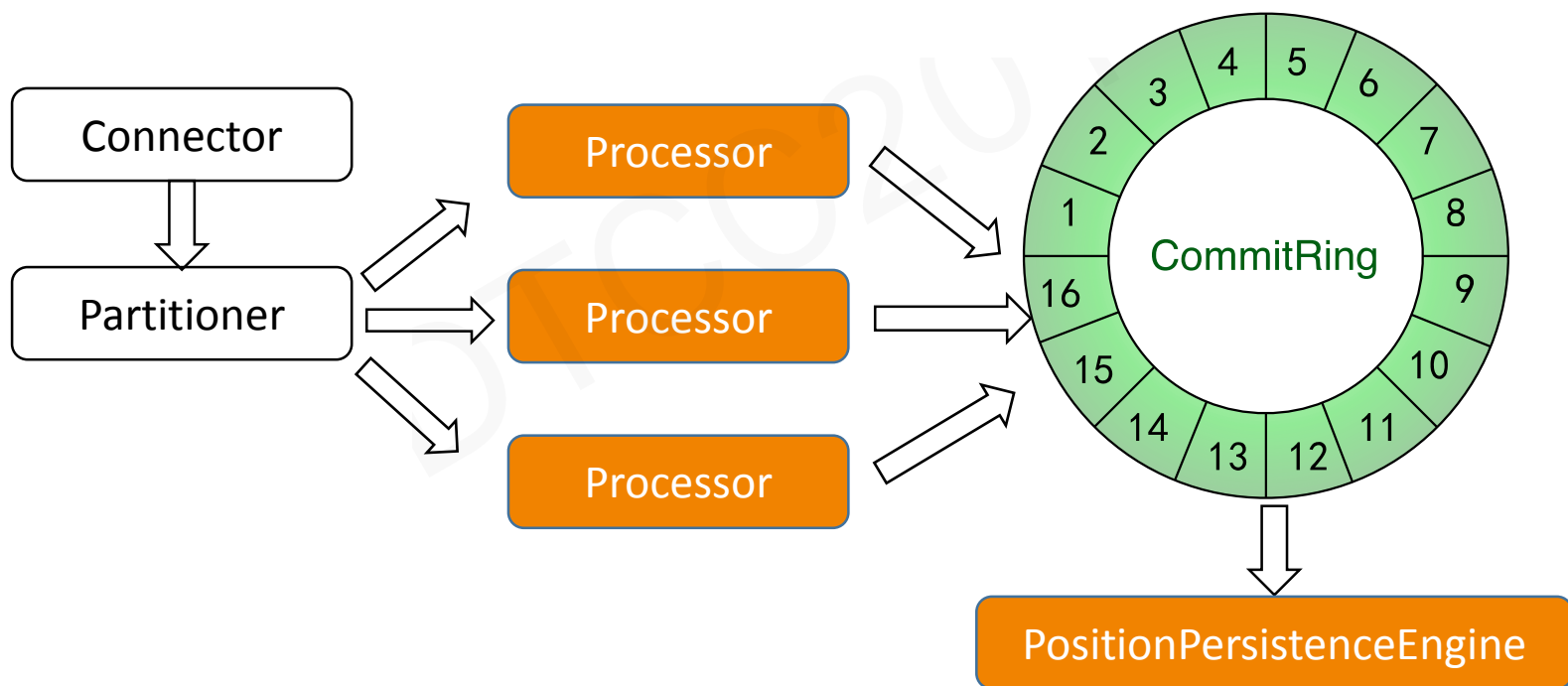
- Binlog中并不记录字段名等相关信息，导致生成的消息只有数据，没有结构。

解决方案：

StreamJob在初次启动时会对mysql中所有的表做一份快照，此后在运行期间当解析到DDL操作时会将原快照取出生成一个新的复本，并在该复本上应用对应的DDL操作，最后生成一个新快照。保证任何时刻的binlog都可以找到其对应的元数据

演进-客户端并行与串行

- 位点使用批量记录的方式，同时要求在并发消费的场景，位点记录时保证该位点之前的所有消息已被正确处理，最终采如下结构：



演进-多场景

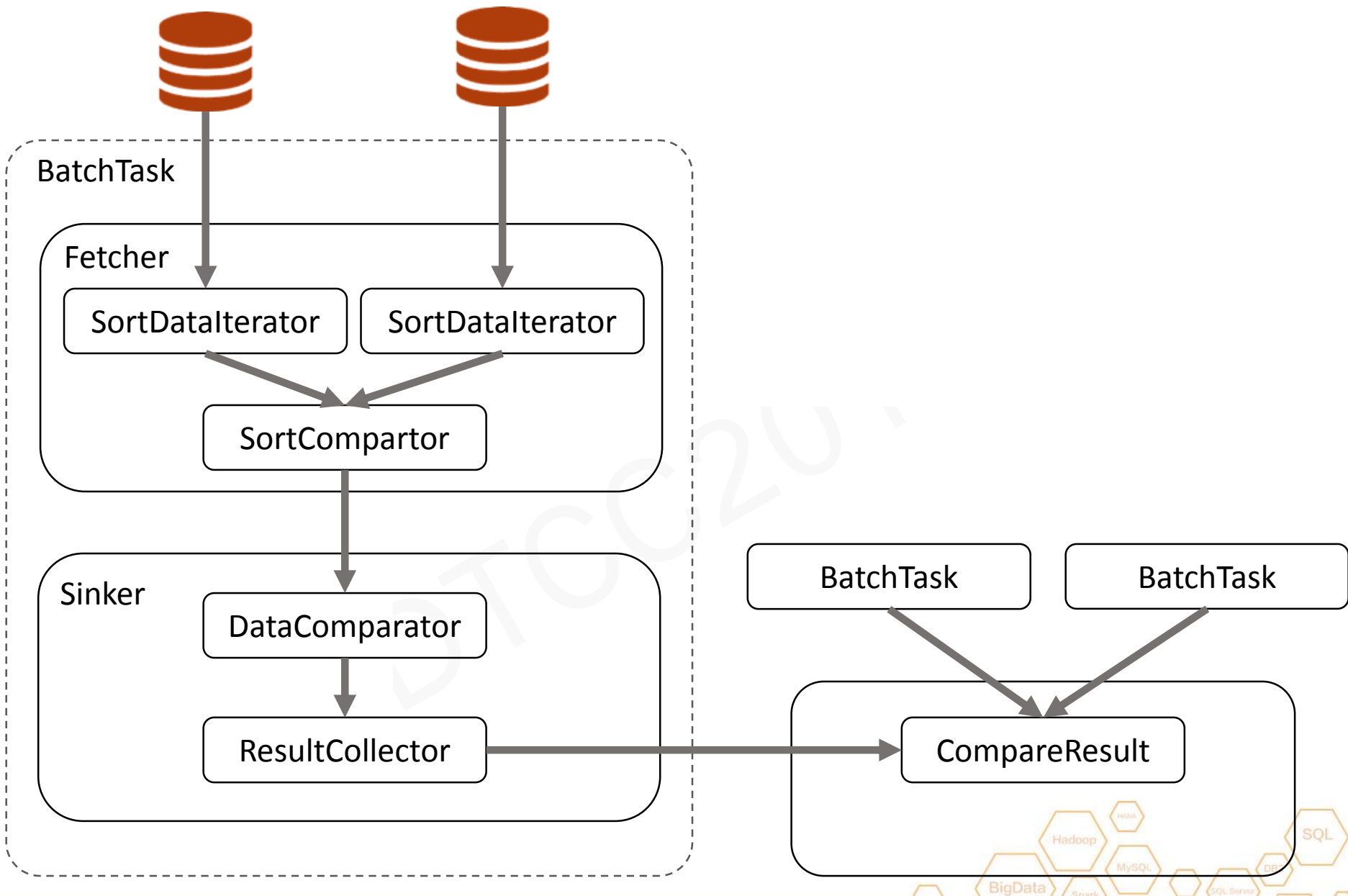
- SQL
- Union
- Join

DTCC2018

演进-一致性校验

- pt_table_checksum
- BatchJob

DTCC2018

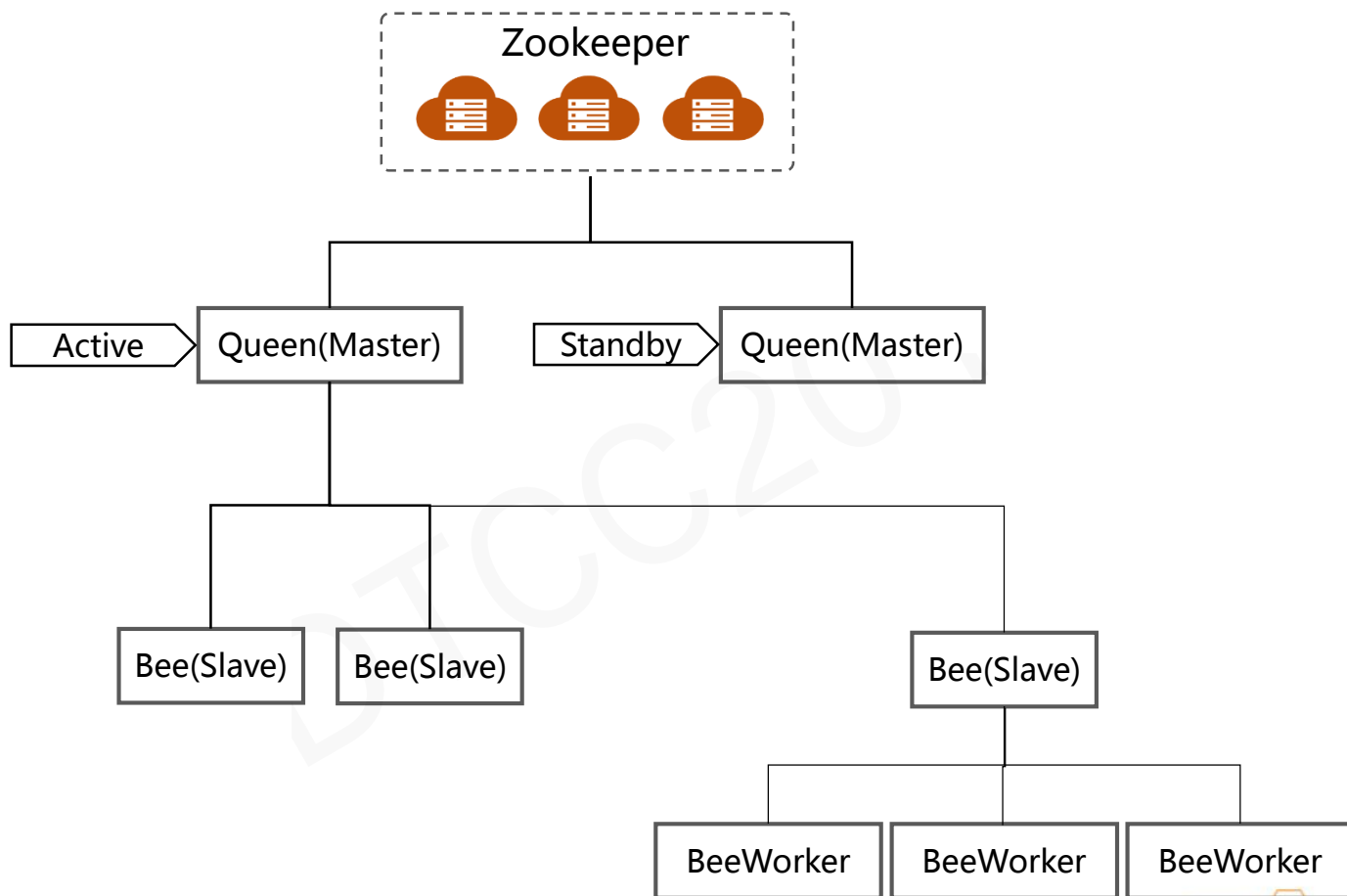


演进-数据修复

- 基于binlog修复
- 直接修复

DTCC2018

演进-资源隔离



演进-易用性

- 监控报警
- 快速接入
- 一键迁移

DTCC2018

CONTENTS

Introduction

□ 背景

□ 演进

□ 架构

□ 现状

现状

- 集群规模
- 接入数据库
- 接入系统

DTCC2018



THANKS





讲师申请

联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动



让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO



联系我们

联系人：黄老师
电话：010-59127187
邮箱：edu@itpub.net
网址：edu.itpub.net
培训微信号：18500940168