



第九届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2018

阿里HBase的数据管道设施实践与演进

孟庆义

DTCC
2018

2018.05.10 – 12 北京国际会议中心



IT168.com

ChinaUnix

ITPUB

自我介绍

- 姓名：孟庆义(花名 天引)
- 阿里巴巴技术专家
- 专注在大数据领域，拥有多年分布式、高并发、大规模系统的研发与实践经验，先后参与hbase、phoenix、lindorm等产品的内核引擎研发，目前负责阿里上万节点的HBase As a Service的发展与落地
- 联系邮箱 qingyi.mqy@alibaba-inc.com

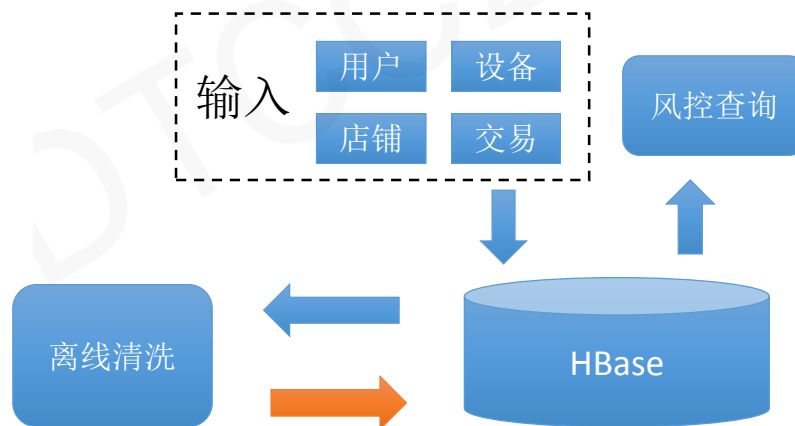
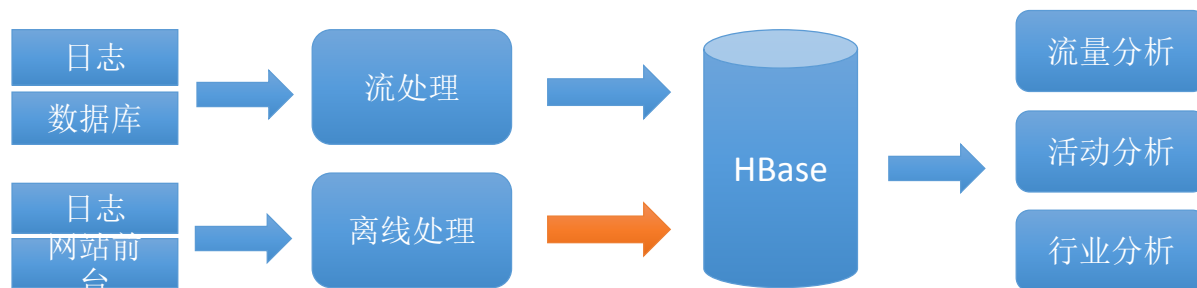
内容大纲

- 数据导入场景
- HBase Bulkload功能
- HImporter系统
- 数据导出场景
- HExporter系统

DTCC2018



数据导入应用场景



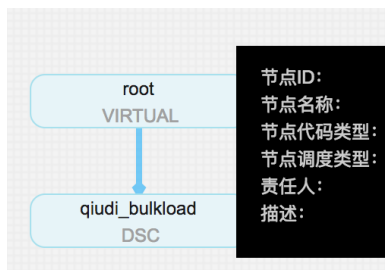
数据导入需要考虑的问题

周期性调度

异构数据源多

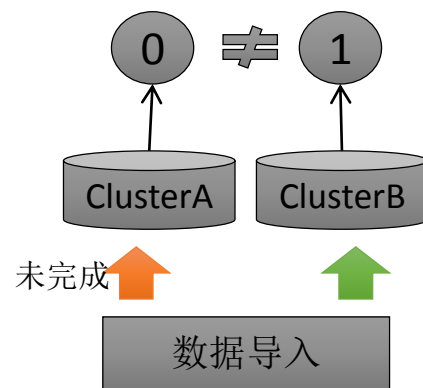
导入效率

多集群下的数据一致性



源类型	MySql
表	MySql Rds for MySql ODPS
增量	Oracle Oceanbase Hbase Postgresql Histore

TB/小时

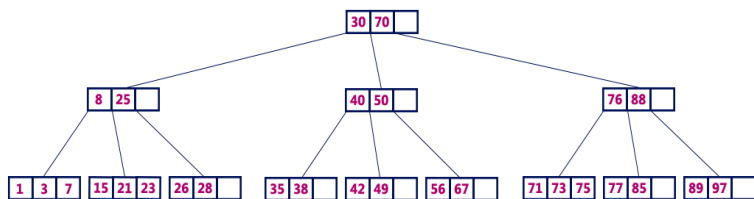


更适合由平台解决

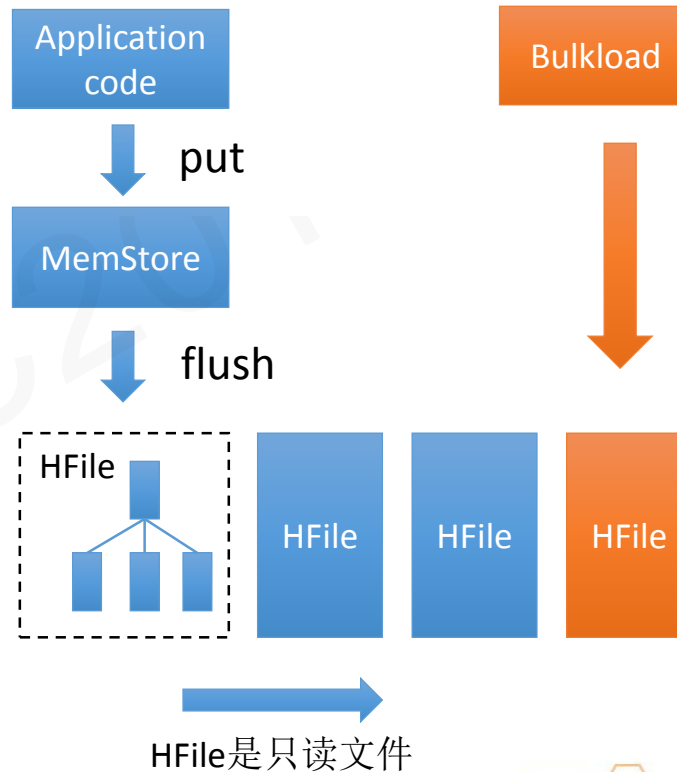
Bulkload介绍

B+Tree

B-Tree of Order 4



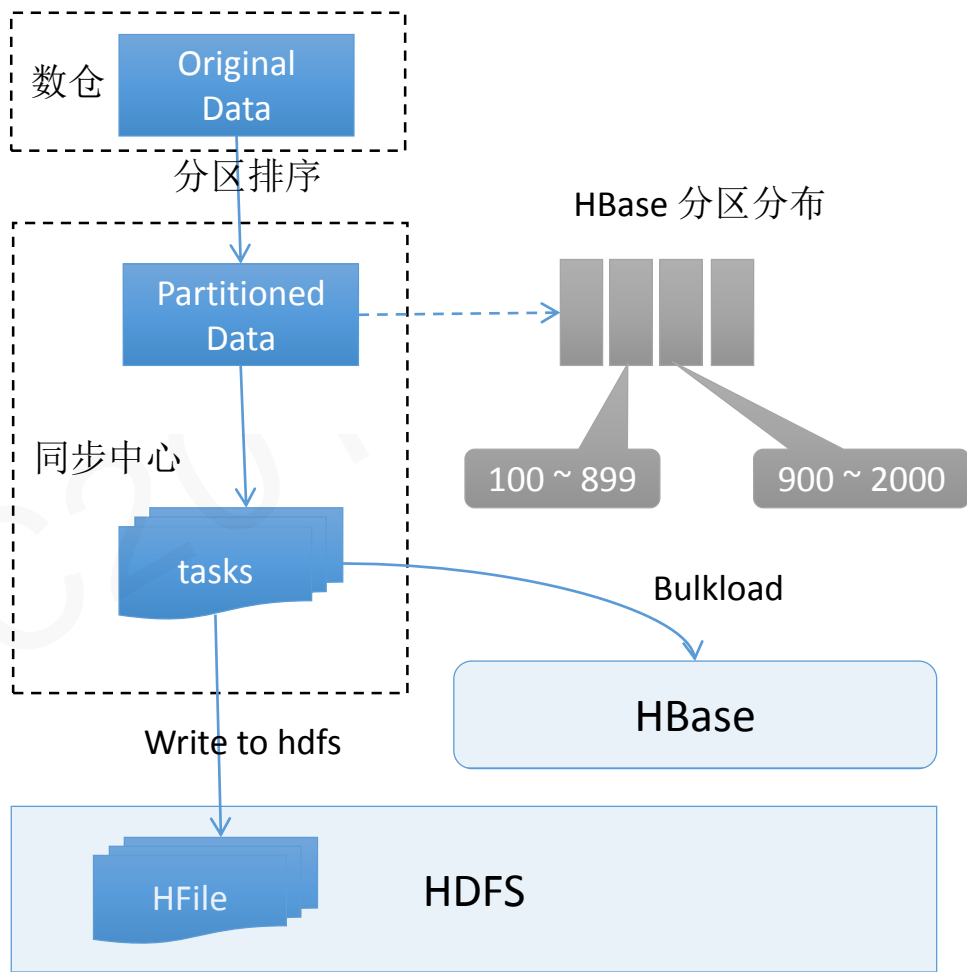
LSM Tree



- Bulkload优势：高吞吐
 - 不需要WAL
 - 避免small compaction
 - 支持离线构建

Bulkload导入

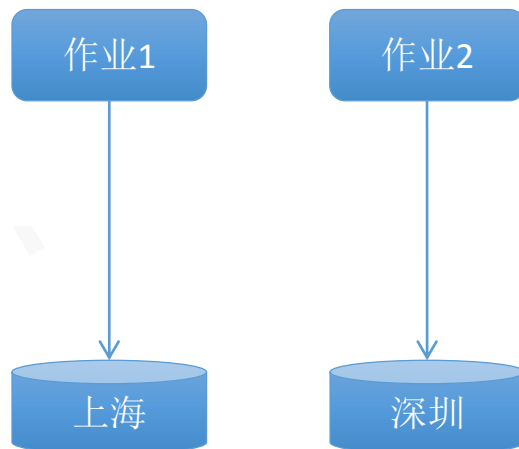
- 导入逻辑作为插件合入同步中心
 - 可调度
 - 适配多种异构数据源
- 按HBase分区对源数据进行分区
 - 避免Split
 - 一个分区对应一个task
 - 一个task一个线程



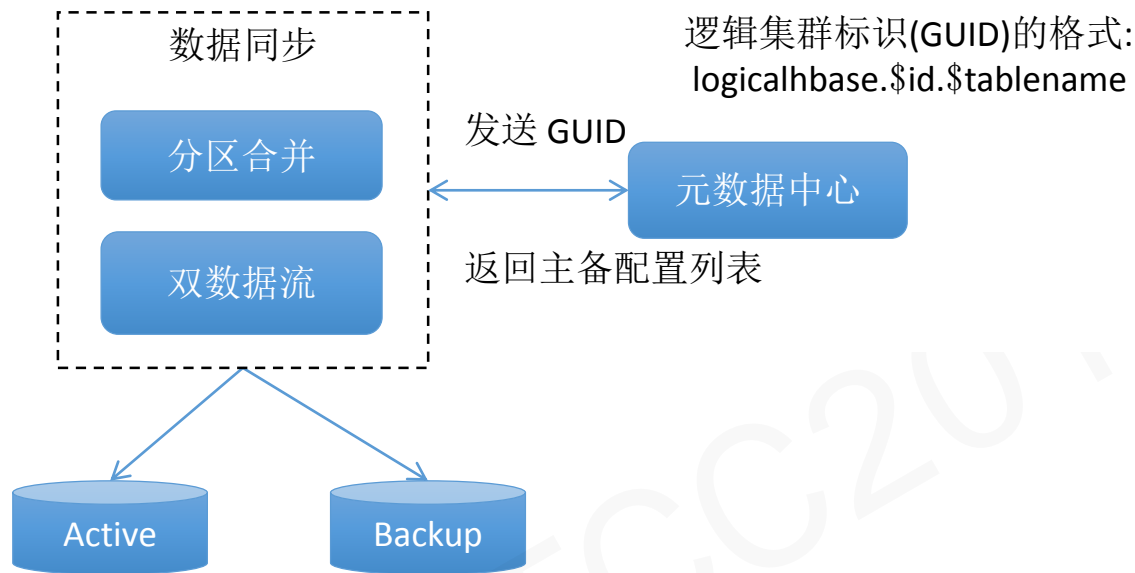
多集群导入

- 很难保证多个任务同时完成，导致一定时间窗口内数据不一致
 - 调度后的运行环境不一致
 - 网络延迟不一致
 - 失败重试
- 集群部署对业务不透明
 - 需要配置多个任务
 - 集群迁移需要重新配置任务

单元化部署模式下的导入



逻辑集群导入



	易用性	一致性	分区排序	编码压缩
多任务模式	重复配置, 不透明	难保证	需要执行两次	两次
逻辑集群模式	配置一次, 迁移无感知	毫秒	执行一次, 但分区数量变多	一次

线上遇到的问题

扩展性



作业单机运行

资源利用率



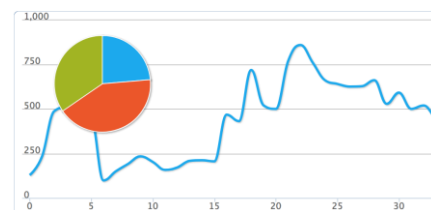
压缩消耗大量CPU

研发效率



快速迭代
在线热修复

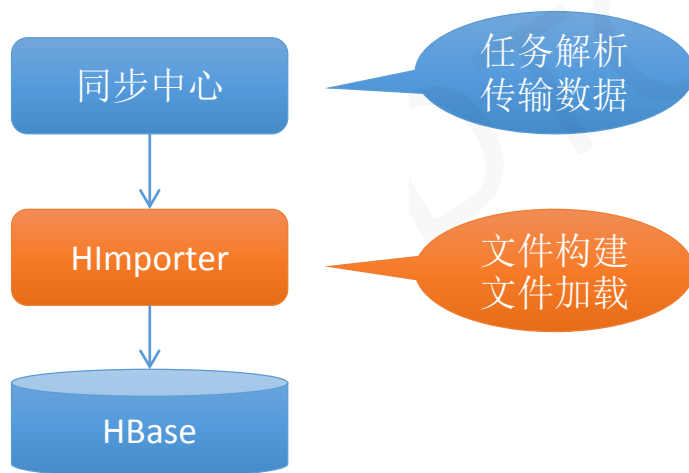
监控



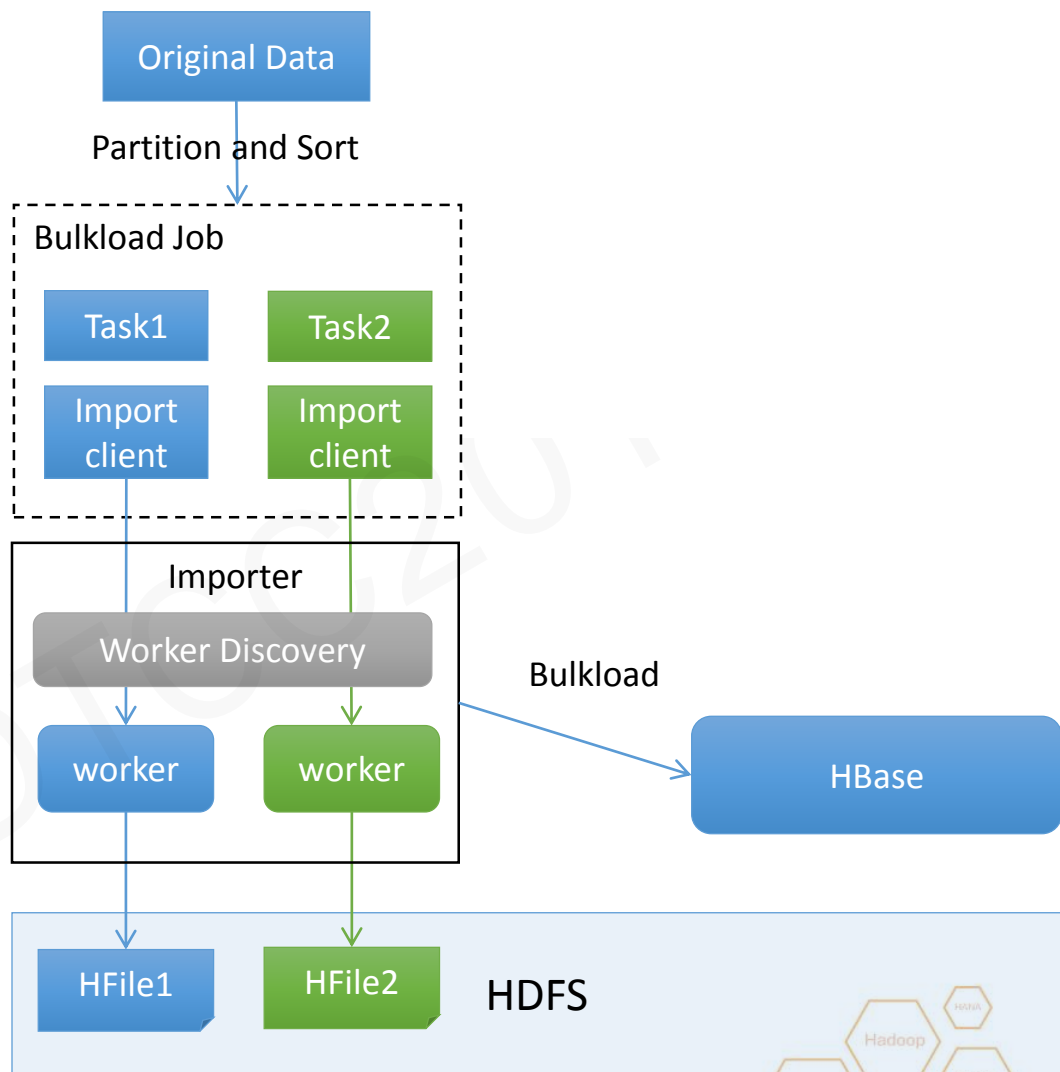
业务统计
性能监控

HImporter

- HImporter是用于辅助数据同步的中间层
- HFile的构建，加载逻辑下沉到HImporter
- 优势
 - 分布式水平扩展，同一个作业的不同任务可以调度到HImporter的不同worker节点
 - 提高资源利用率，将压缩等CPU密集操作下降到HImporter
 - 快速迭代，HImporter的运维和迭代与同步中心独立
 - 独立监控，HImporter可按照自己的需求实现监控

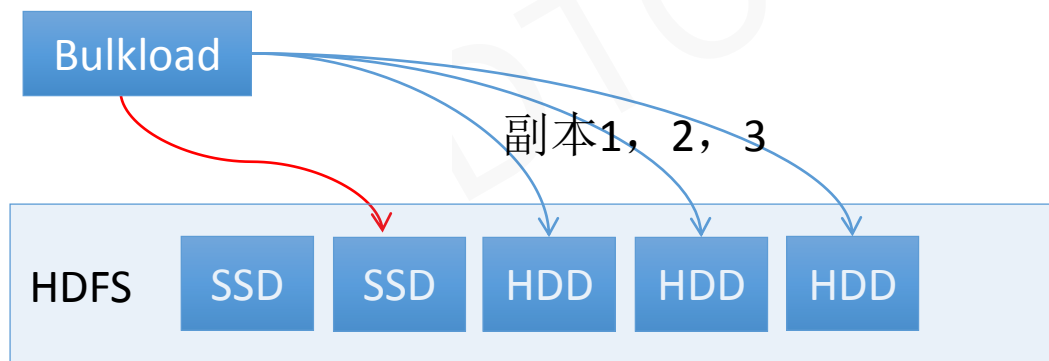
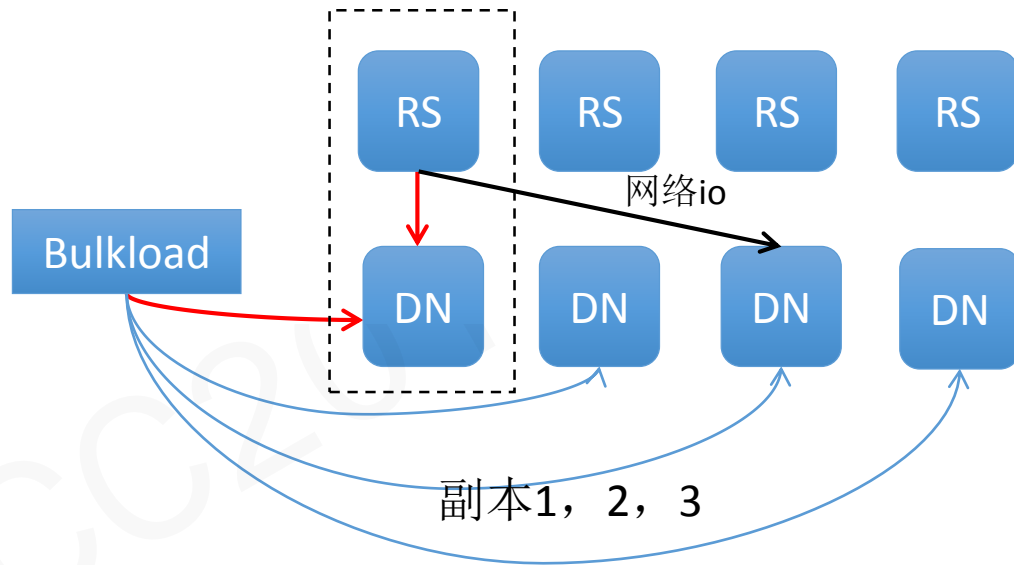


HImporter



提高Bulkload稳定性

- 支持本地化率
 - HDFS默认随机放置
 - 提高本地化率可减少网络IO
- 感知存储类型
 - HDFS默认选择同一介质
 - 感知表存储类型
 - 支持混合存储



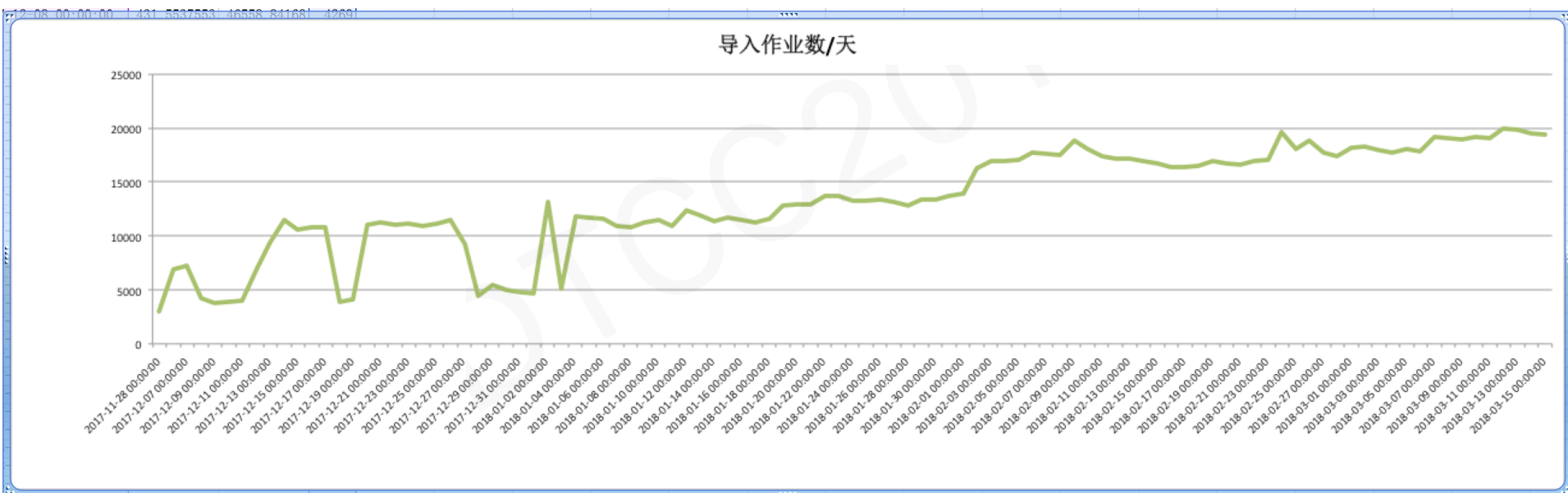
HImporter 功能迭代

- 表属性感知
 - 混合存储
 - 新压缩编码
 - 表级别副本数
- 保证本地化率
 - 将Hfile写入到分区所在服务器，保证本地化率和存储特性，对一些rt敏感的业务效果明显
- 支持轻量计算
 - 支持MD5，字符串拼接等函数
- 安全隔离
 - 避免对外暴露HDFS地址
- 支持Task级别重试



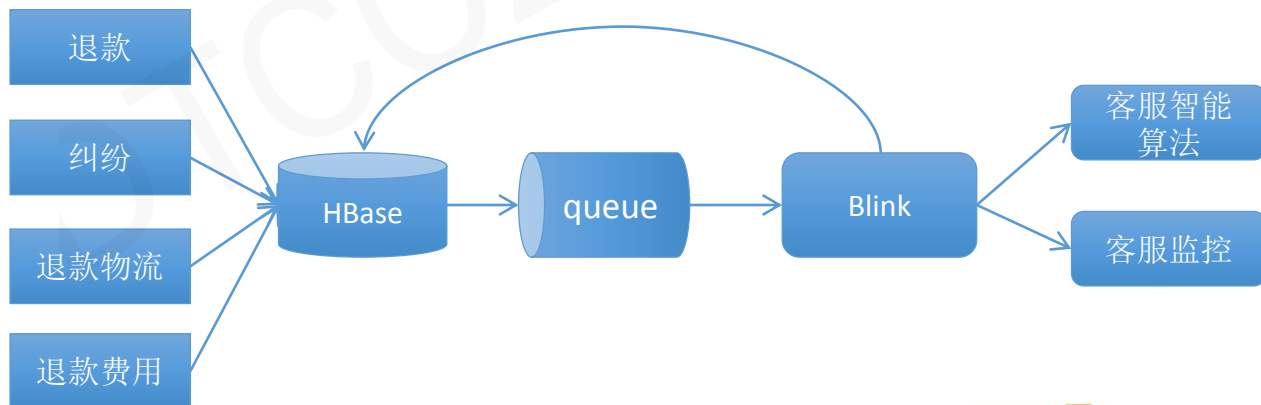
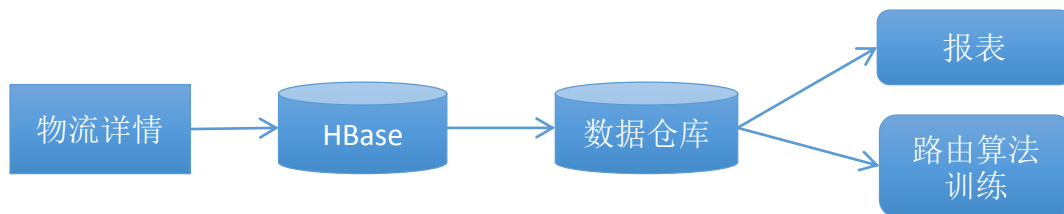
Hlmporater 业务统计

- 作业数 2万/天
- 导入数据量 1.5PB/天



增量数据导出场景

CAI
N!ΔO
菜鸟联盟



DTCC
2018

数领先机 智赢未来 (9)

IT168.com

ChinaUnix

ITPUB

增量数据导出需要考虑的问题

离线数据的T+1处理特点

吞吐量

实时性

主备流量切换

离线数据分区

00:00~01:00 01:00~02:00

p1

p2

.....

GB/秒

实时消费 毫秒
T+1消费 分钟

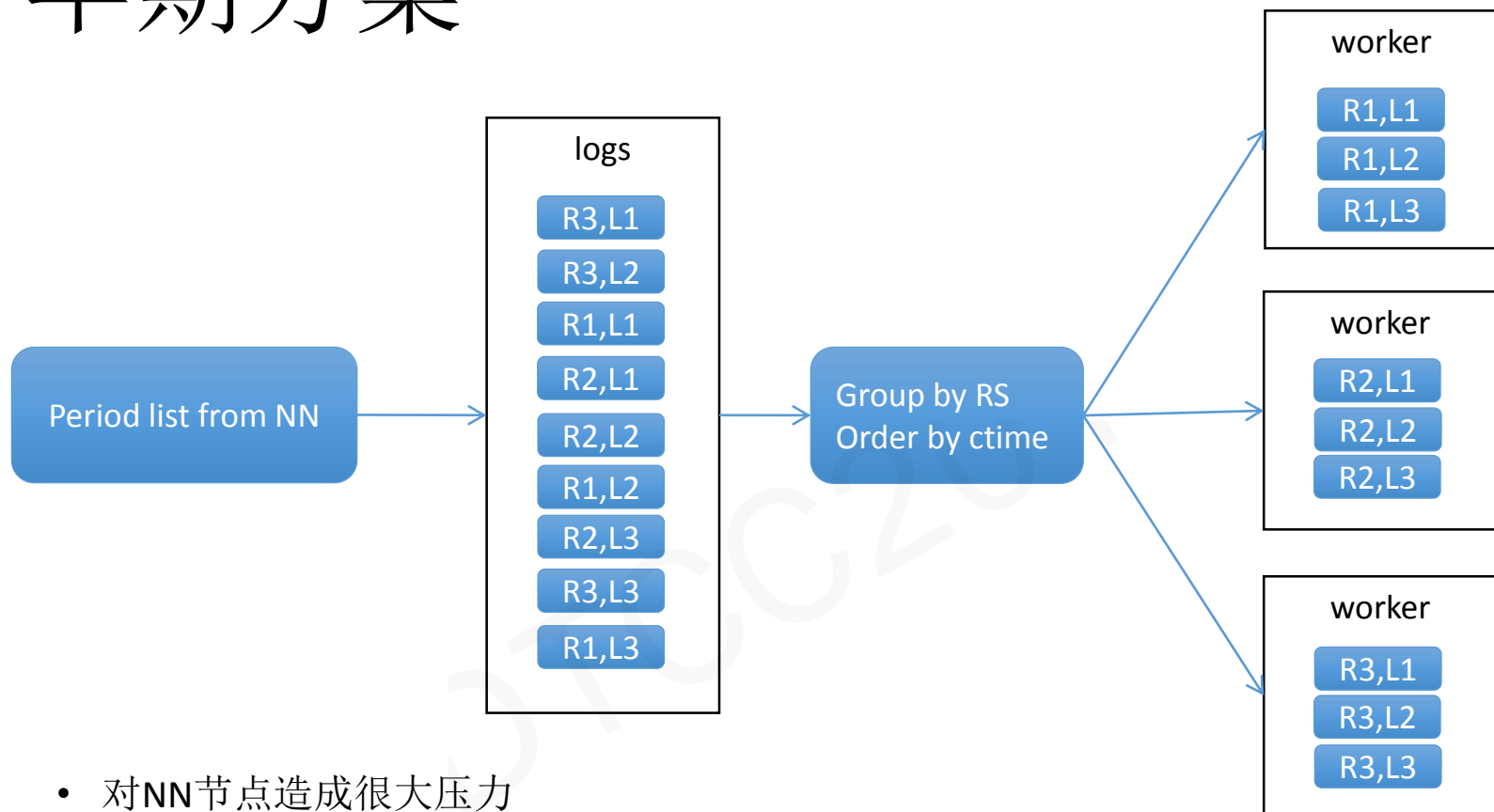
Active

如何采集?

Export

Backup

早期方案

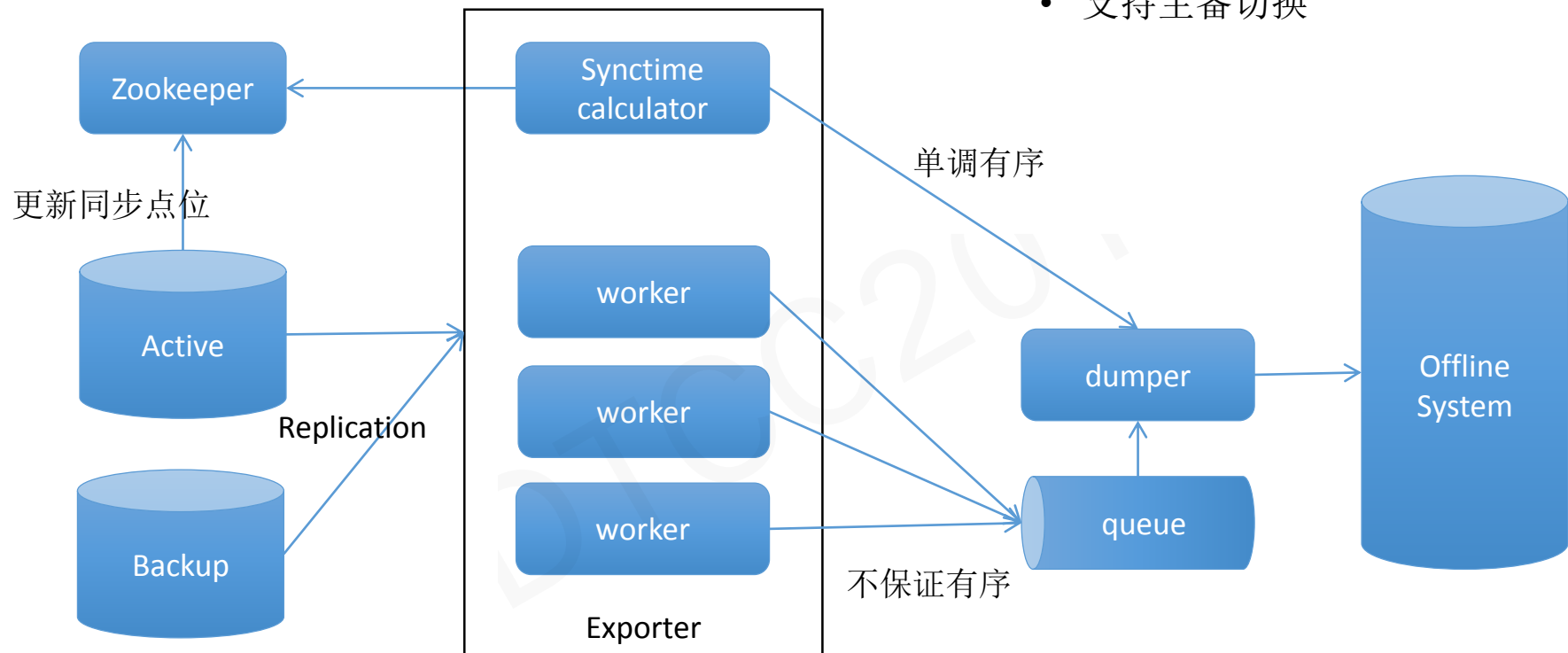


- 对NN节点造成很大压力
- 无法应对主备切换
- 日志热点处理能力低



HExporter1.0

- 无状态 Worker
- 支持主备切换



HExporter1.0优势

- 主备流量切换不影响数据导出
 - 识别数据来源，过滤非原始数据
- 独立的同步时间流
 - 保障数据按有序时间分区Dump到数据仓库
- 复用HBase replication框架
 - 降低开发工作量
 - 复用HBase的监控，运维体系

DTCC2018

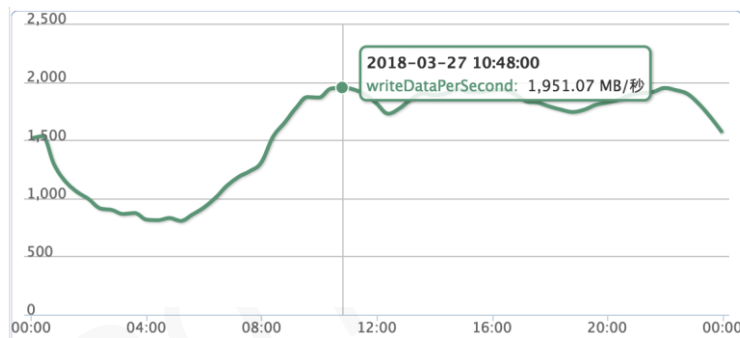
HExporter1.0 优化

- 减少拓扑网络中的数据发送
 - 备库避免向Exporter发送重复数据
- 远程辅助消化器
 - 空闲的机器帮助消化热点
- 避免发送小包
 - HExporter在接收到小包后，等待一段时间再处理
- 同步通道配置隔离
 - 实时消费链路和离线消费链路可以采用不同的配置
- 数据发送前压缩

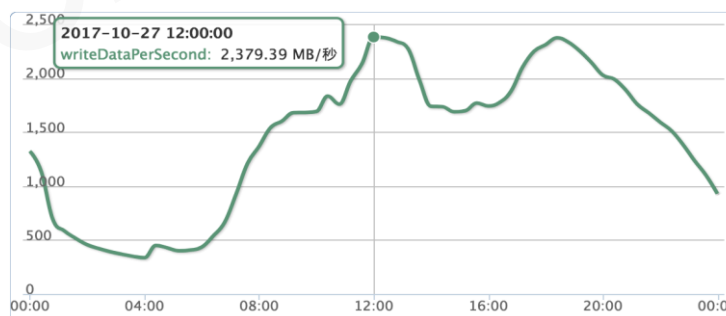
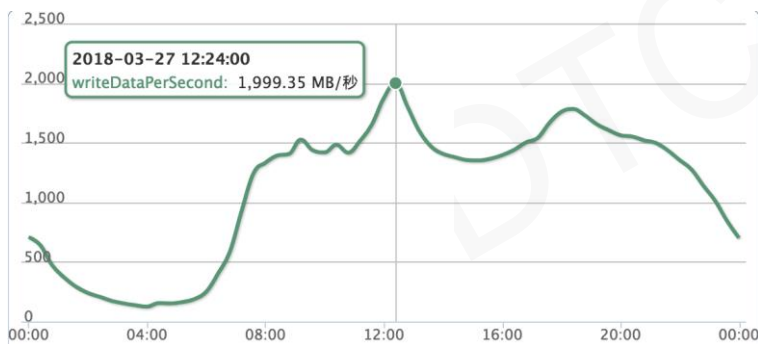
HExporter1.0 业务统计

- 采集集群数 57个
- 采集任务数 554个
- 日常峰值 6GB/s

集团业务



蚂蚁金服业务

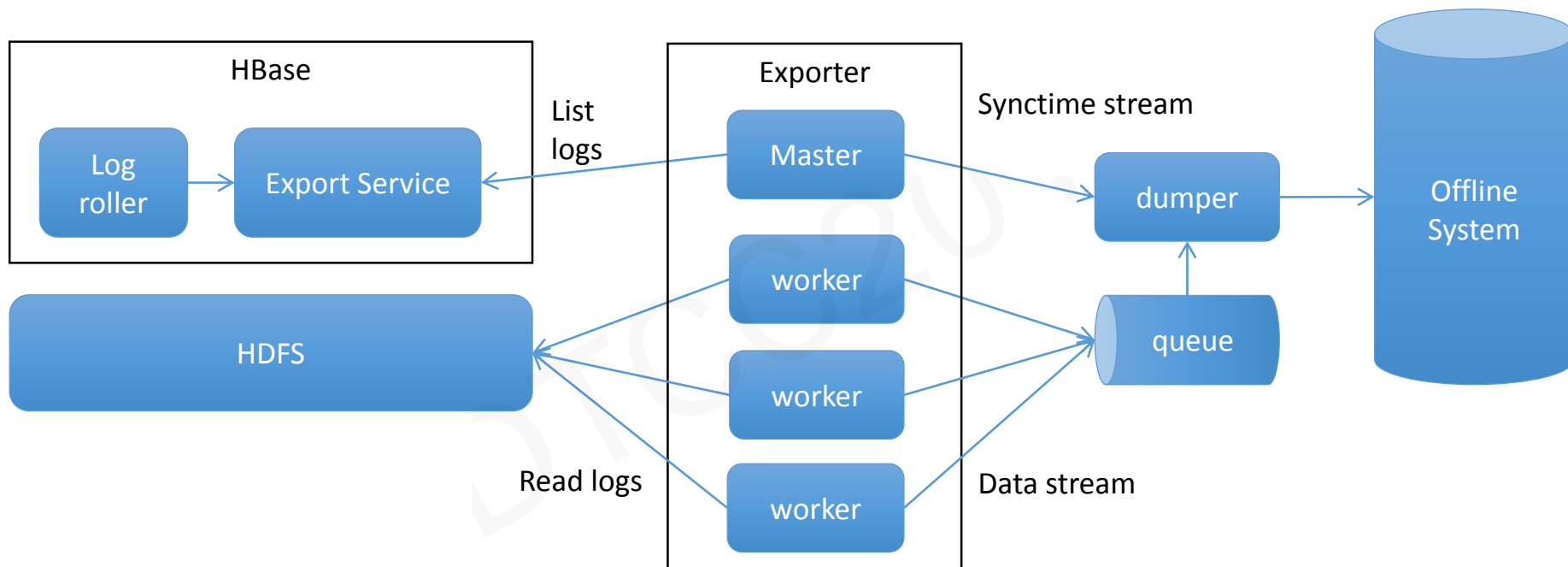


HExporter1.0问题

- 举例：午餐和晚餐时间，业务写入流量产生高峰，离线同步延迟
 - 数据的发送与在线写入争抢资源，发送赶不上写入
 - 找空闲机器辅助？不一定有
 - 扩容？资源浪费，周期长
 - 思路：如果离线同步的资源可以和在线资源隔离，利用离线大池子可以削峰填谷
- 开发一个功能用了2天，升级HBase集群用了2周
 - 升级慢的原因是Hbase是有状态的，升级过程不能影响线上业务稳定性
 - HBase集群数量多
 - 思路：Exporter的worker是无状态的，如果把所有逻辑都放在Exporter，那么升级、扩容会简单快速



HExporter2.0

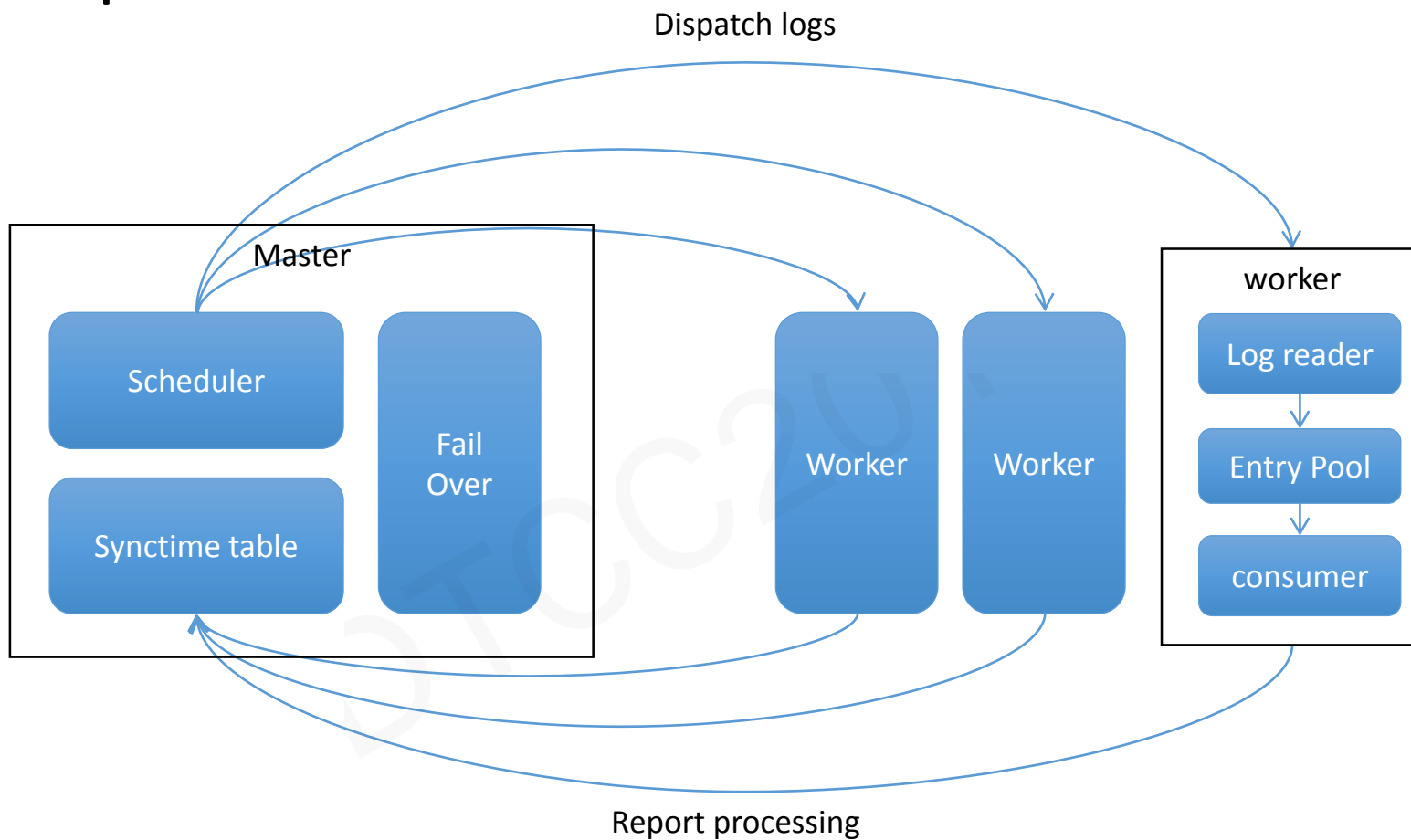


Export Service

- Consumer
 - Register/Unregister
 - List logs
- Log life cycle
 - Create
 - State change (Writing => Closed => Finished)
 - Log clean
- Log Entry
 - Filesystem, Path
 - ClusterID, ServerName, CreateTime, ...



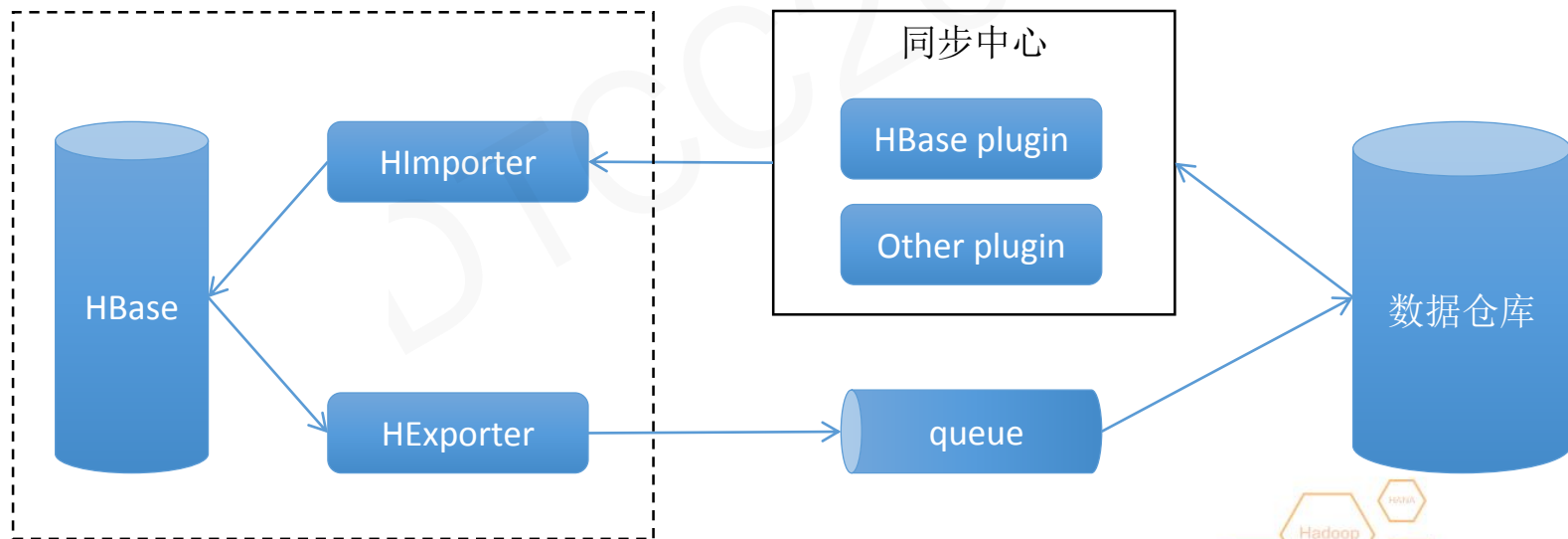
HExporter 2.0



AliHBase数据通道

– 中间层(HImporter,HExporter)核心价值

- 易扩展 分布式水平扩展
- 高可靠 自主识别主备切换，封装对HBase访问
- 迭代快 架构解耦，快速迭代
- 稳定 无状态，节点对等



欢迎加入HBase中文社区

- HBase中文技术社区 <http://www.hbase.group/>

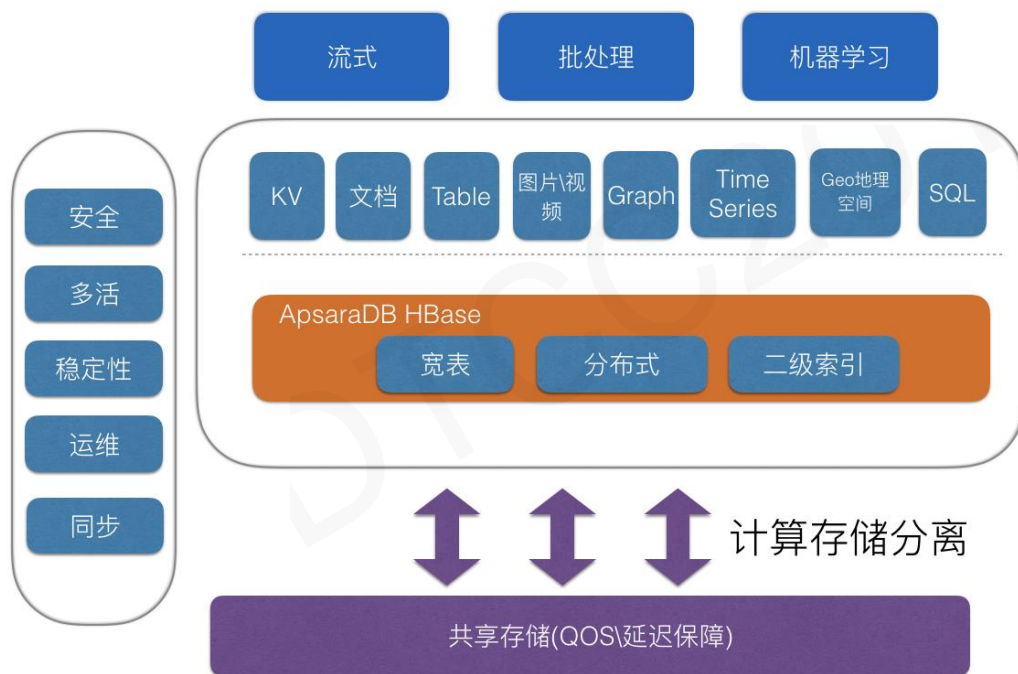


技术社区微信公众号



钉钉技术交流群

云数据库 HBase 版（ApsaraDB for HBase）是基于 Hadoop 且100%兼容HBase协议的高性能、可弹性伸缩、面向列的分布式数据库，轻松支持PB级大数据存储，满足千万级QPS高吞吐随机读写场景



多模式数据库

kv\document\table\graph\sql
ts\geo地理空间

存储计算无感知扩容

存储：200g ~ 10P
计算：1000 ~ 5000w QPS

存储计算分离

存储可以单独扩容
解决计算存储阻抗
成本较低

完全兼容开源

求才若渴，欢迎加入我们！杭州，北京，深圳

qingyi.mqy@alibaba-inc.com





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