Druid在滴滴的应用实践与平台化建设

刘博宇

目录

Druid特性介绍-Druid是什么？

Druid是针对时间序列数据提供低延时的数据写入以及快速交互式查询的分布式OLAP数据库。

Druid特性介绍-时序数据库

TSDB(Time-series database)

- 时间序列数据
- 低延时写入
- 快速聚合查询

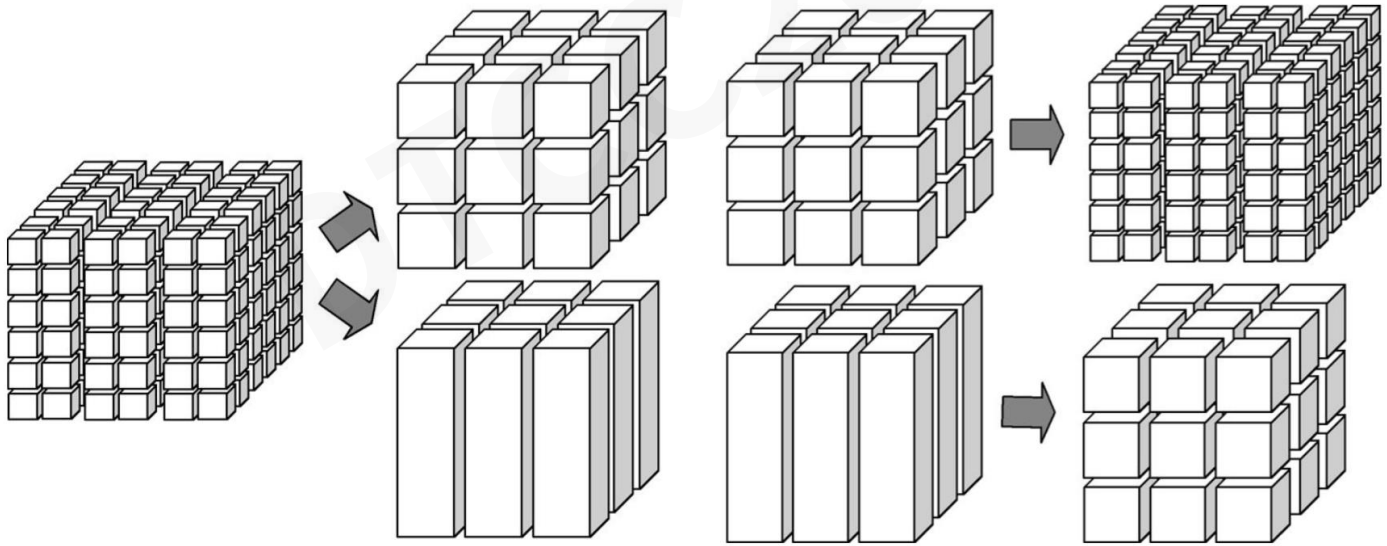
典型的TSDB：InfluxDB、Graphite、OpenTSDB

- 写入即可查 – 内存增量索引
- 下采样，RDD – 预聚合
- Schema less – 需要预先定义schema

Druid特性介绍-OLAP数据库

OLAP数据库 - 上卷、切块、切片、下钻等操作

- 数据检索引擎 – ES
- 预计算 + kv存储 - Kylin
- SQL on Hadoop – Presto、SparkSQL



Druid特性介绍-OLAP数据库

数据检索引擎 – ES

① 结构化数据与非结构化数据，明细查询与聚合能力

② 存储空间开销大

③ 数据的写入与聚合开销大

Druid 结构化数据 & 预聚合

① 结构化数据 较弱的明细查询能力

② 存储空间更小

③ 针对数据的写入与聚合进行优化

Druid特性介绍-OLAP数据库

预计算 + kv存储 - Kylin

KV存储通过预计算来实现聚合，key涵盖了查询参数，值就是查询结果

① 查询速度极快

② 损失了查询的灵活性，复杂的场景下，预计算过程可能十分耗时

③ 只有前缀拼配一种索引方式，大数据量下复杂过滤条件性能下降

④ 缺少聚合下推的能力

Druid 列式存储 & Bitmap索引

① 查询速度不如KV存储

② 内存增量索引，增量预聚合，写入即可查

③ 任意维度列组合过滤、聚合，查询灵活

④ Scatter & Gather模式，支持一定的聚合下推

Druid特性介绍-OLAP数据库

SQL on Hadoop

① SQL支持强大

② 无冗余数据，不需要预处理

③ 分钟级响应

④ QPS低

Druid

① SQL支持有限

② 必须预先定义维度指标

③ 亚秒级响应

④ 高并发

Druid在滴滴的应用-使用概况

规模

- 多个集群数百台机器
- **千亿级**日原始数据写入量
- **TB级**日落盘数据量
- 数百实时数据源，千级实时写入任务
- **近千万级**日查询量

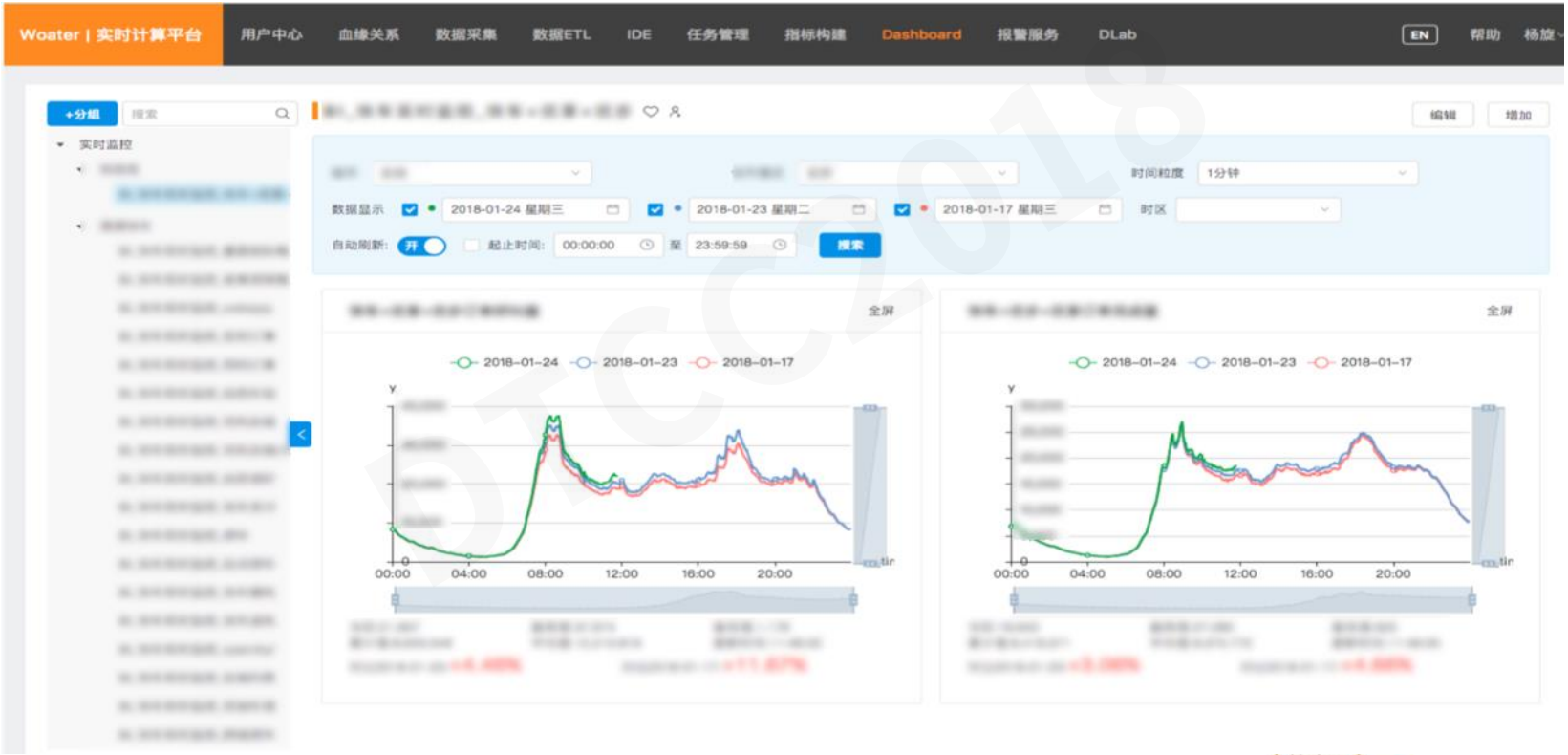
承接业务

- 监控、实时报表、大屏展示等业务

Druid在滴滴的应用-应用案例

业务实时监控

承接公司所有核心业务的指标监控与告警



Druid在滴滴的应用-应用案例

实时报表类应用

运营数据分析、客户端网络性能分析、客服应答实时统计等等

Druid在滴滴的应用-应用案例

大屏展示类应用

客服服务状态大屏

Druid平台化建设

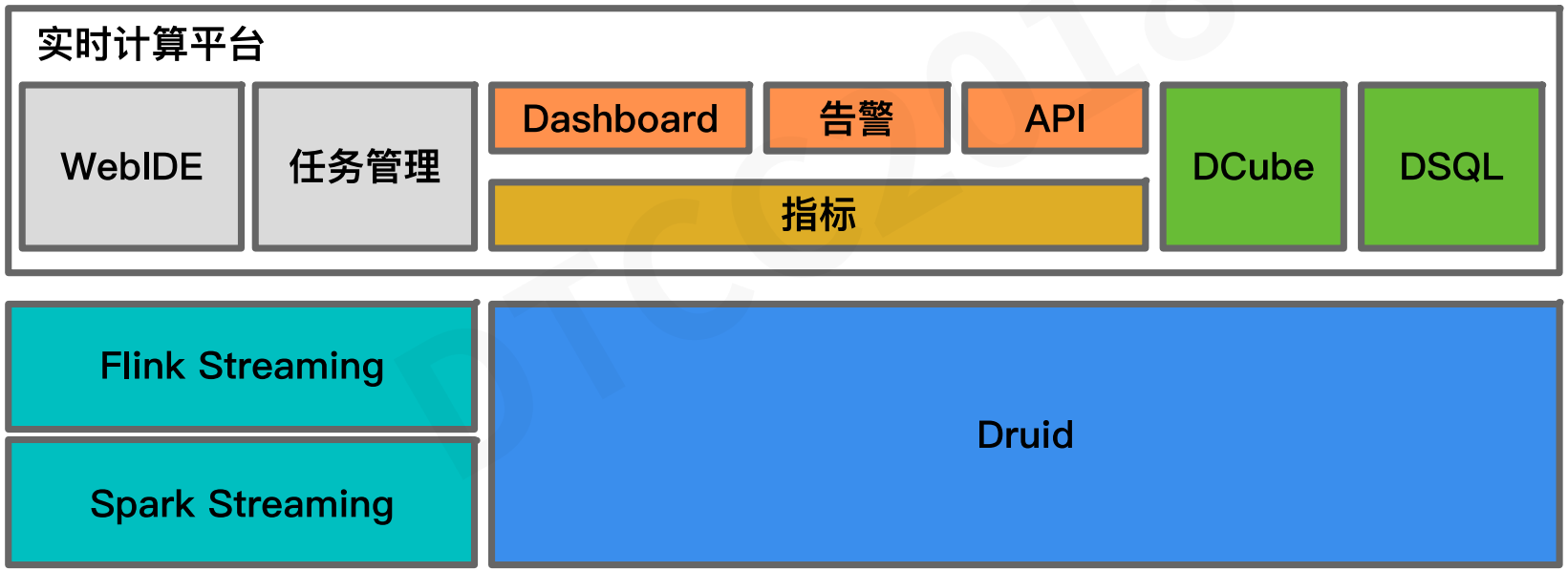
背景

- 业务数据主要来源，日志、binlog
- 公司统一数据通道Kafka
- 业务监控指标多样，逻辑复杂多变
- Druid接入配置较复杂，工单接入方式成本高
- 数据进入Druid之前通常需要流计算处理
- 数据链路较长，上下游关系需要梳理
- Druid服务需要提供数据可视化能力

Druid平台化建设

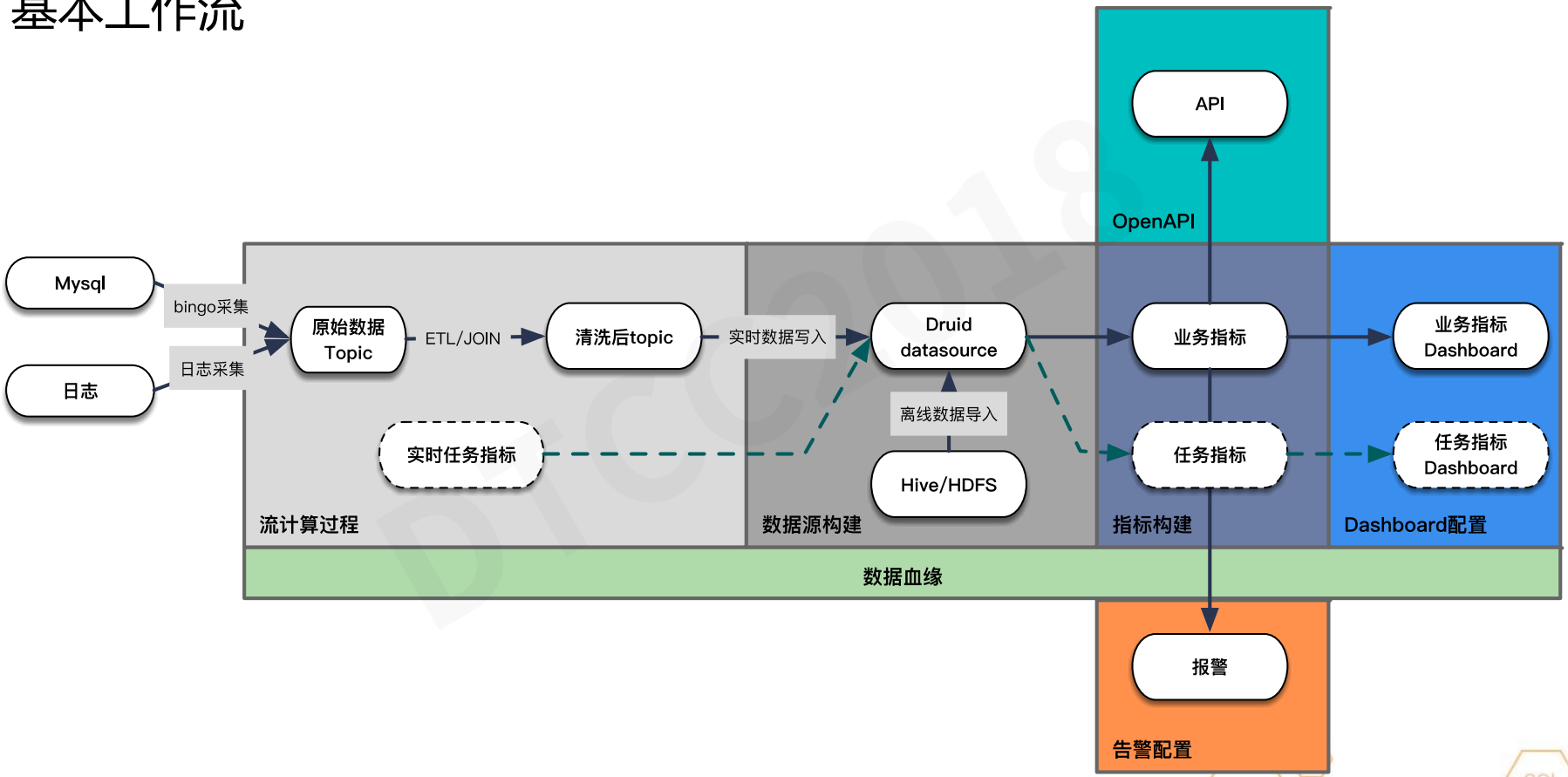
实时计算平台

提供流计算，Druid数据储存，指标查询，数据可视化一站式服务。



Druid平台化建设

基本工作流

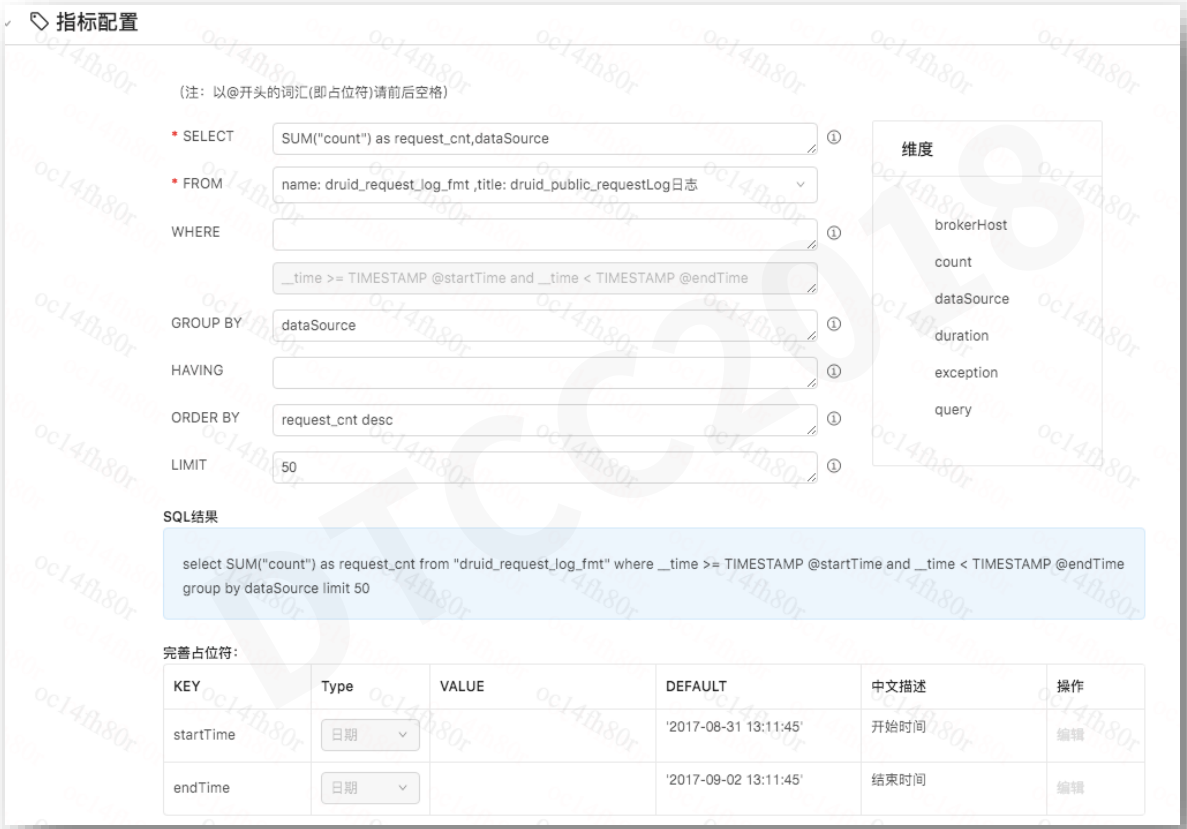


Druid平台化建设

Druid数据源用户自助接入

Druid平台化建设

Druid查询Web化配置，100%SQL



Druid平台化建设-稳定性

挑战

1. 核心业务与非核心业务共享资源，存在风险。
2. 用户提交任务配置、查询不合理，造成异常状况，甚至影响集群稳定性。
3. 随着业务的快速发展，Druid依赖组件热迁移到独立部署环境。

Druid平台化建设-稳定性

针对不同重要程度的业务共享资源的问题

1. Druid集群异地双活，核心数据源集群级双活
2. 统一网关建设

① 对用户屏蔽多集群细节

② 根据用户身份进行查询路由，实现查询资源隔离

3. 业务分级：

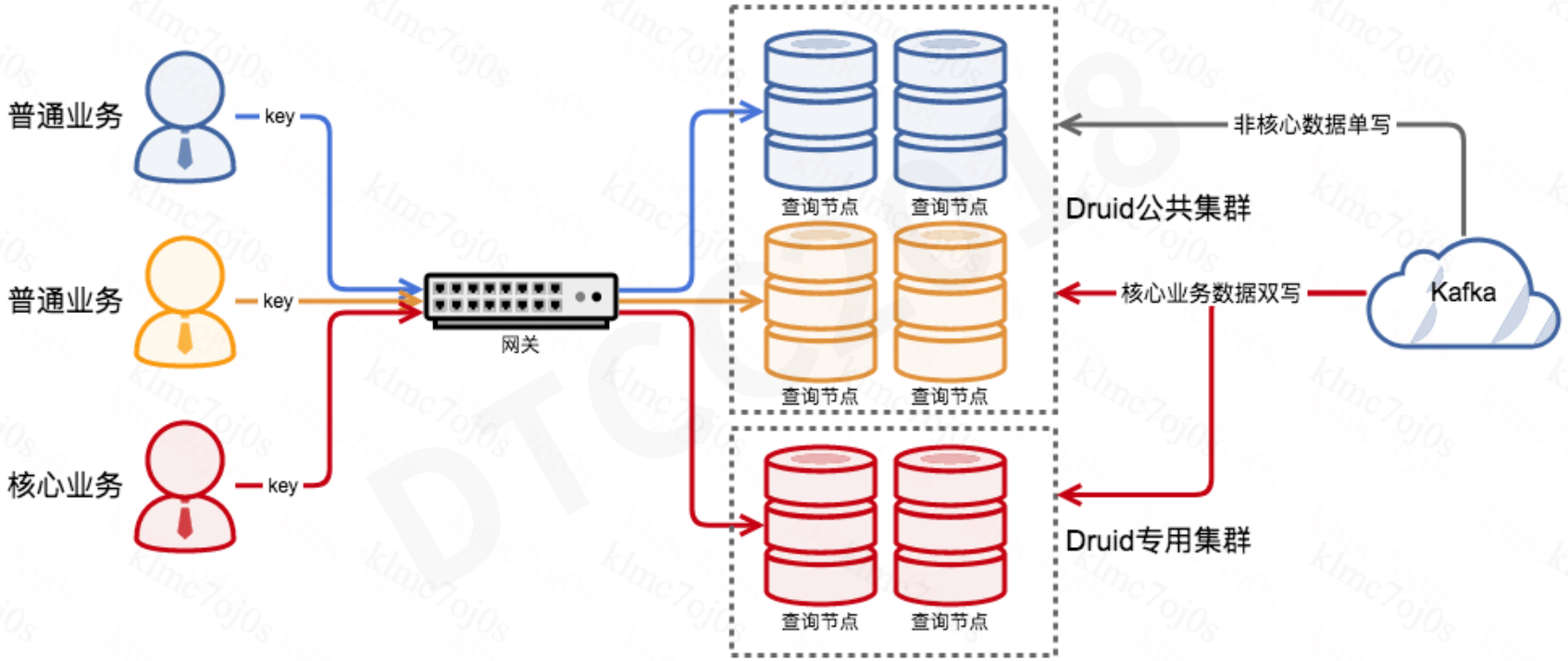
① 核心业务集群级双活；

② 对查询资源需求较高的大业务分配独立查询资源组

③ 其他使用默认资源池

Druid平台化建设-稳定性

异地双活、业务分级、资源隔离



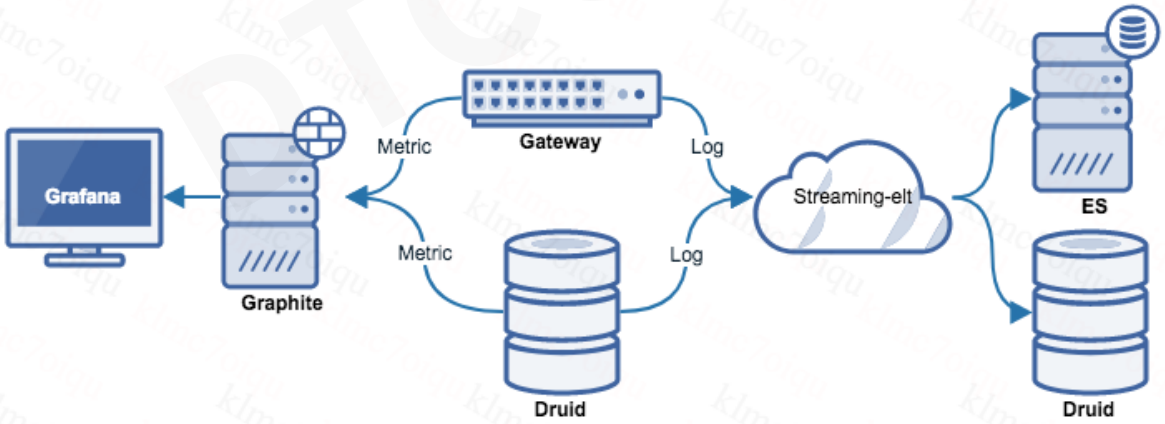
Druid平台化建设-稳定性

针对用户配置与查询不合理造成的异常

1. 引擎层面bad case防范（earlyMessageRejectPeriod的case）
2. 封装druid原生API，提供更合理的默认配置项
3. 完善指标监控体系与异常定位手段，确保能捕捉到异常查询

日志与指标收集

结合Druid的聚合查询能力与ES的明细查询能力进行问题定位

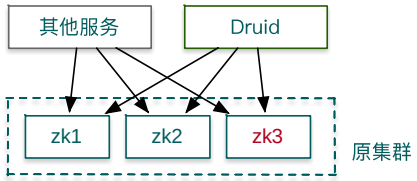


Druid平台化建设-稳定性

第三方依赖热迁移

Zookeeper迁移：扩容-集群分裂-缩容的迁移方案

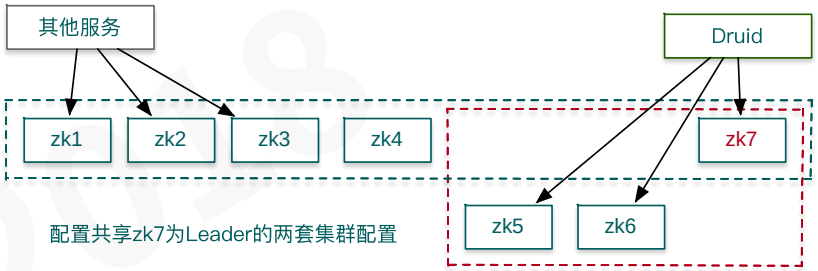
1.



2.

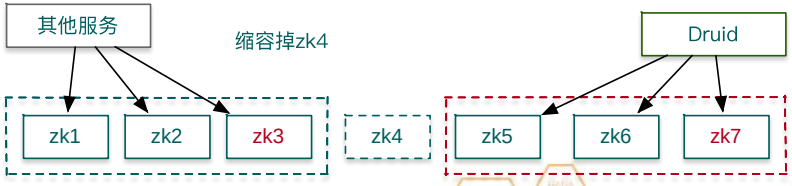
3.

4.



5.

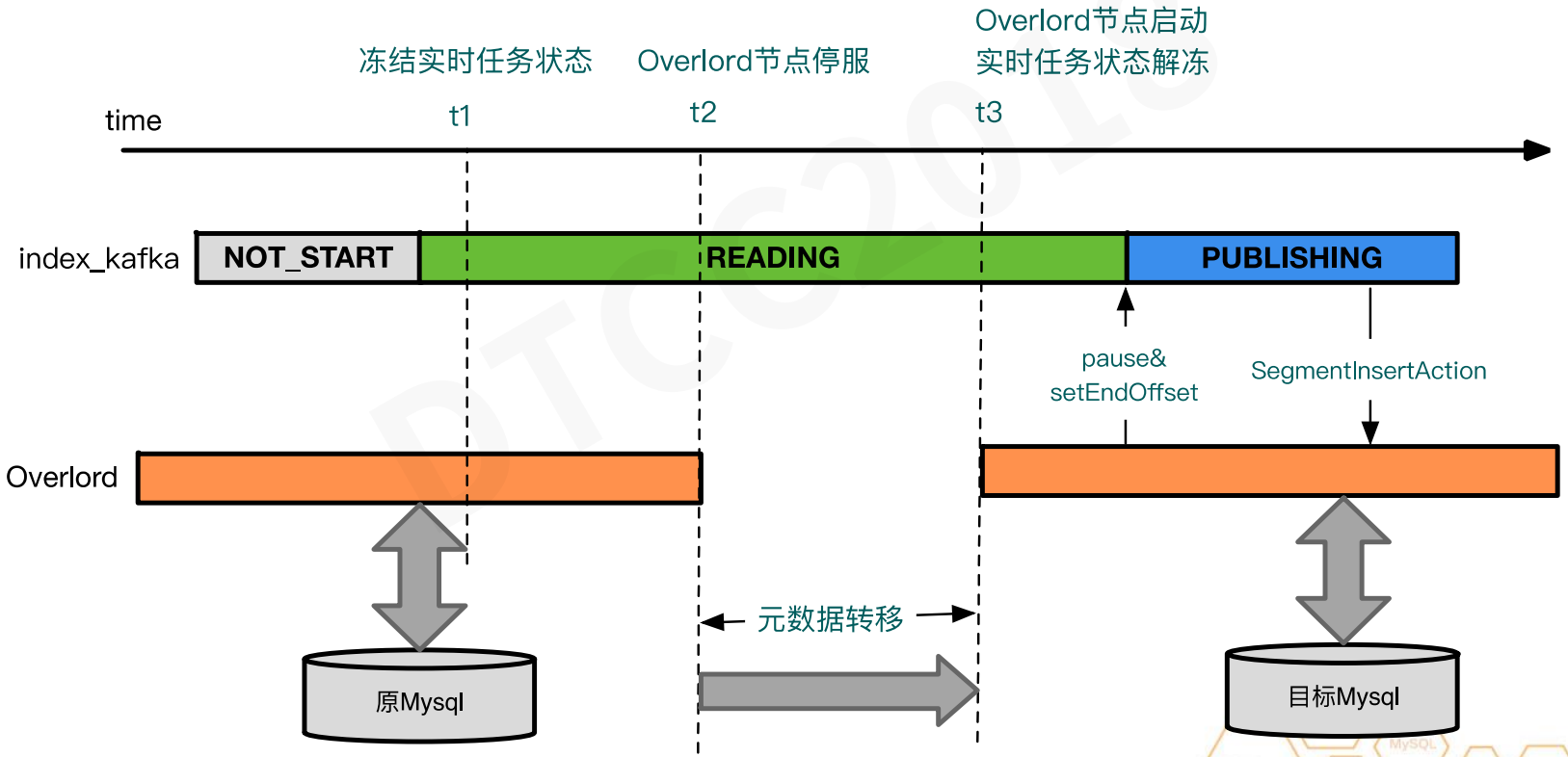
6.



Druid平台化建设-稳定性

第三方依赖热迁移

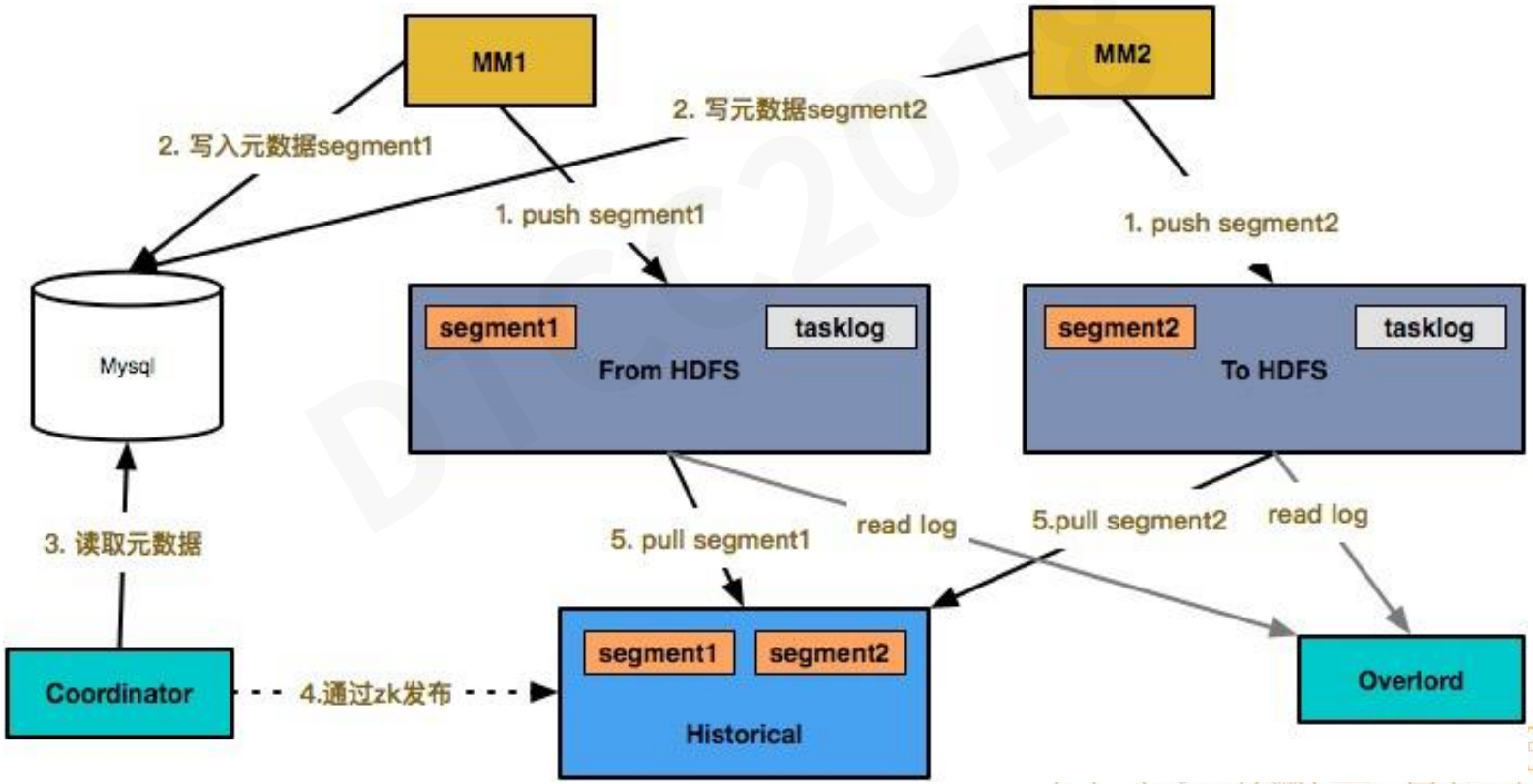
MySQL迁移：开发实时任务状态冻结API（针对Kafka-indexing-service），保证元数据的不变性，随后进行迁移



Druid平台化建设-稳定性

第三方依赖热迁移

HDFS迁移：保证历史节点可以读取两个HDFS集群，混动升级MM，改变增量数据写入地址；批量修改元数据，改变存量数据的加载地址。



Druid平台化建设-性能优化

实时数据接入方式对比

Standalone Realtime Node

- 数据消费任务为单机模式，任务失败后无法恢复
- 使用Kafka高阶API，多任务消费数据时，难以保证副本任务消费相同的数据

Tranquility + indexing-service

- 任务失败后无法恢复，如果所有副本任务都失败，那么还是会丢失数据
- 数据迟到容忍窗口与任务时长挂钩，无法做到容忍较长时间的数据迟到

Kafka-indexing-service

- 实时任务数据消费依赖Overlord服务，所以Overlord单机性能将会成为集群规模的瓶颈
- 由于Segment与Kafka topic的partition关联，容易造成元数据过度膨胀，引发性能问题

Druid平台化建设-性能优化

问题背景：

主要Kafka-indexing-service作为数据写入方式，具有高可用、接入便捷的优势，但是高度依赖Overlord节点服务。 Overlord节点高峰期的性能瓶颈导致：

① Druid消费能力下降

② 实时任务调度不及时，实时任务状态判断错误

经过定位，瓶颈有以下原因

① Mysql查询性能问题

② 元数据JSON化存储，反序列化耗时

③ ZK watch回调单线程模型事件处理排队

Druid平台化建设-性能优化

针对Mysql查询瓶颈

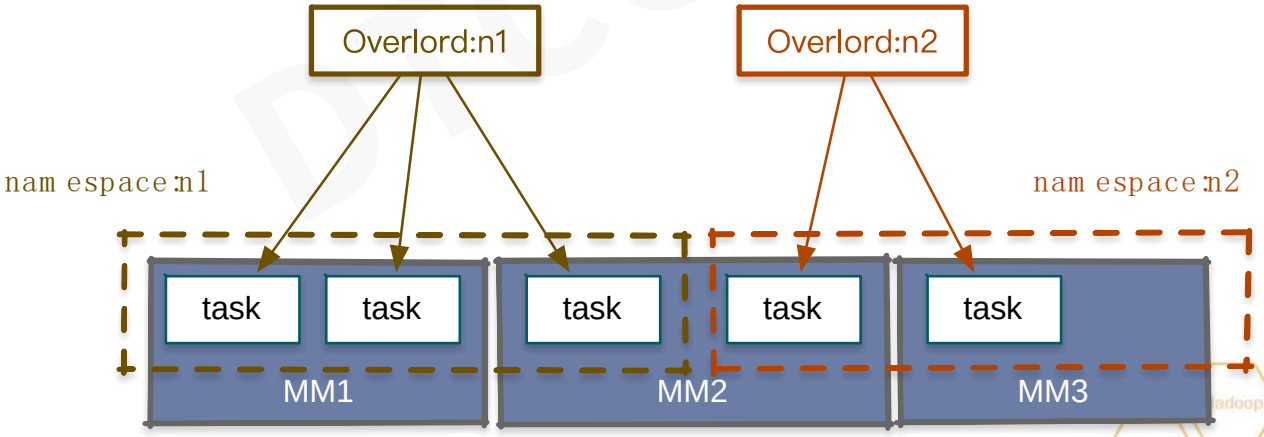
① Druid元数据存储索引优化

② 元数据合并精简，Segment定时Merge，合理设置数据生命周期

③ 数据库连接池DBCP2参数修改

针对反序列化与Watch回调问题

对Druid进行多Overlord改造，引入namespace概念，增加Overlord水平扩展能力



展望

- Druid数据消费能力依赖kafka topic的partition，引入Flink等流计算引擎，提升单partition消费能力，解耦对topic partition的依赖。
- Overlord大量服务需要涉及对Mysql的直接操作，单机性能瓶颈仍存在，后续将会对高并发服务进行内存化改造。
- Coordinator任务处理单线程模型需要优化。
- On-yarn，提升资源利用率，简化运维。

THANKS


HANA
Hadoop
SQL
MySQL
DB2
BigData
Spark
SQL Server
NewSQL
Oracle
MPP
NoSQL
OLAP

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

微学堂

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