苏宁物流实时大数据的探索与实践

邢建垒

V3.0

目录

天眼系统简介

基本功能

- 1. 面向物流及售后领域
- 2. 实时数据监控
- 3. 多维度数据分析

业务场景

- 1. 物流订单全链路实时跟踪
- 2. 仓储作业监控
- 3. 包裹分拨监控
- 4. 车辆线路的实时监控

5s目标

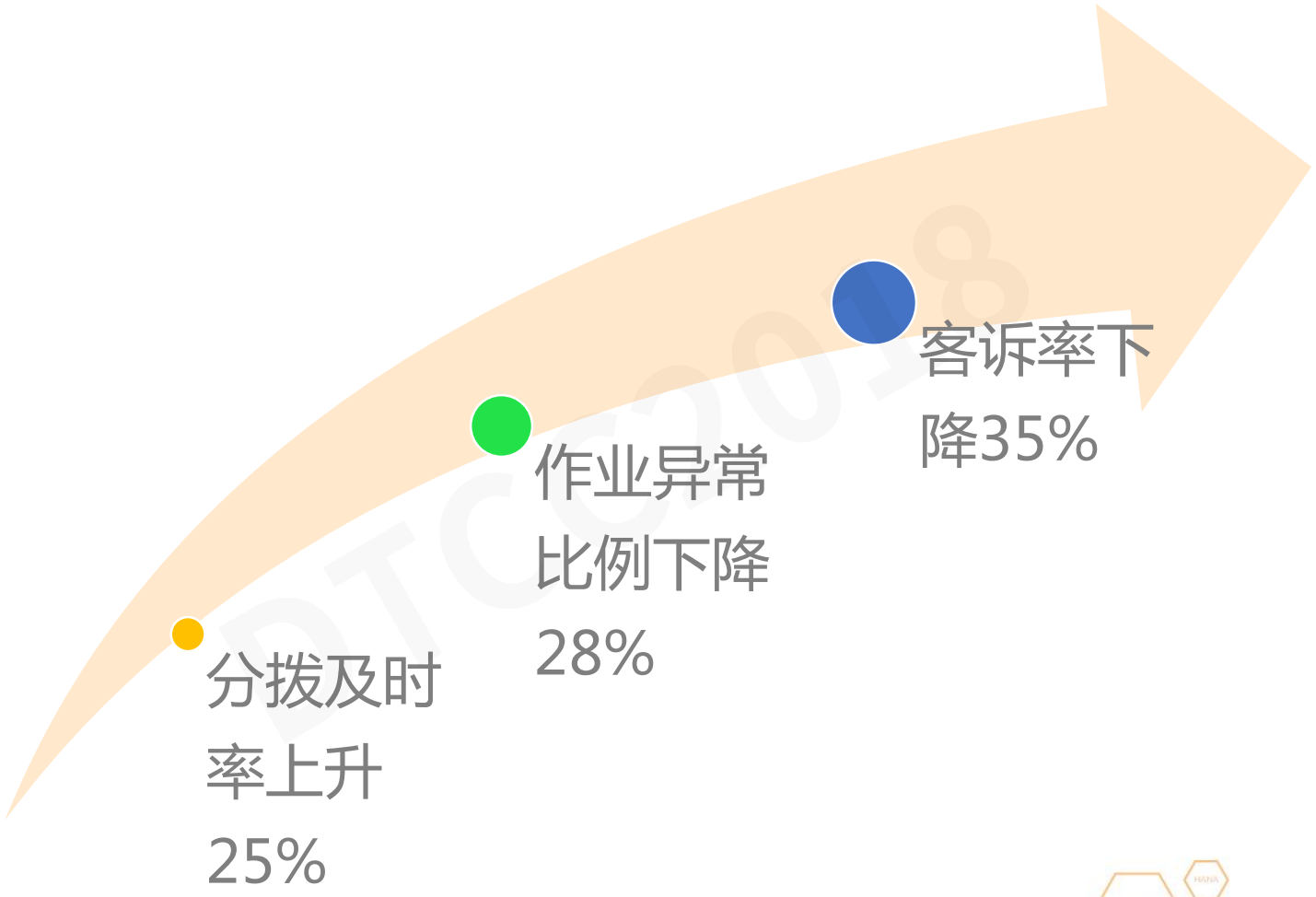
天眼系统指标

每天可以接收订单状态10亿条。

每秒可处理订单数30万条。

核心监控报表分钟级数据延迟

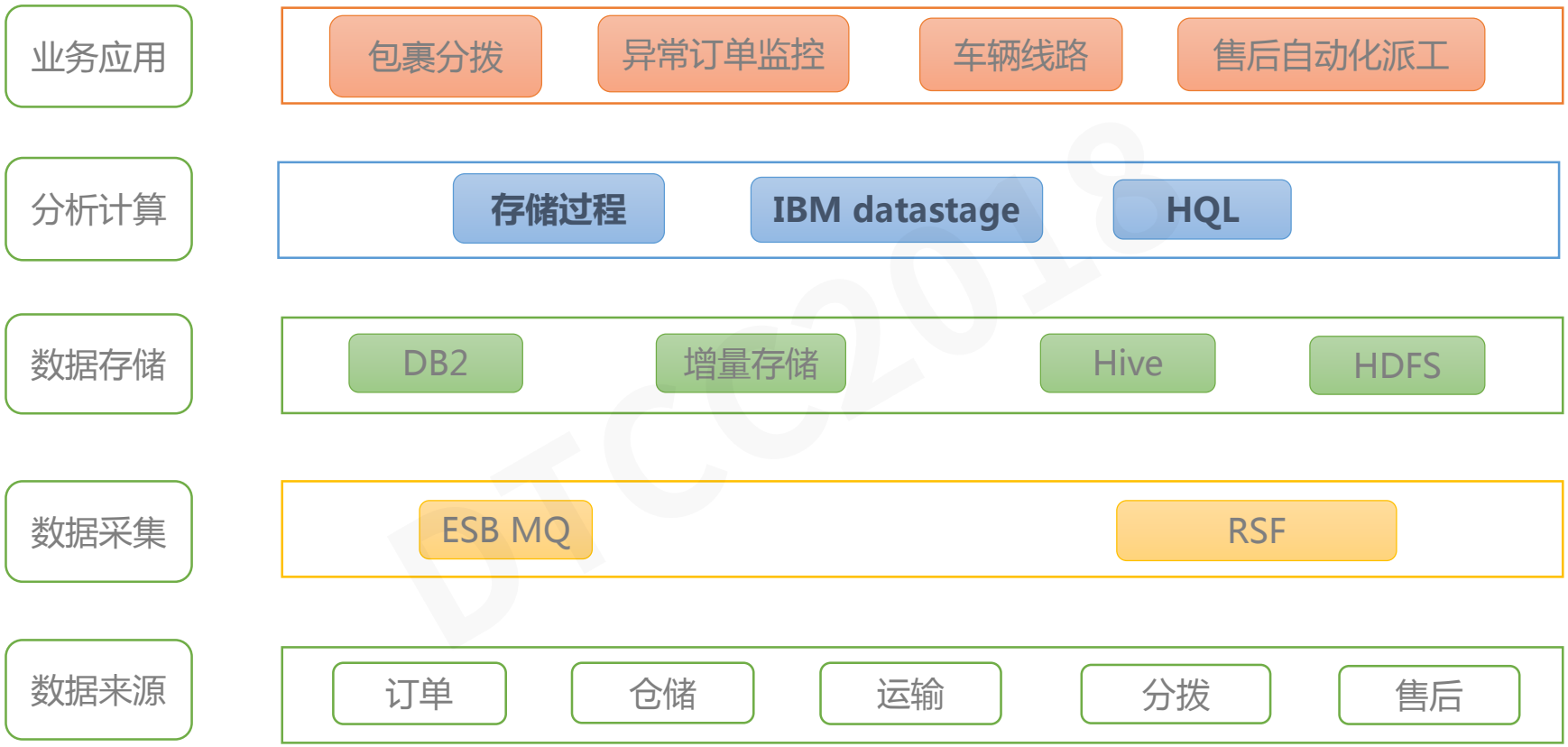
天眼实施效果



目录

苏宁实时技术架构演进

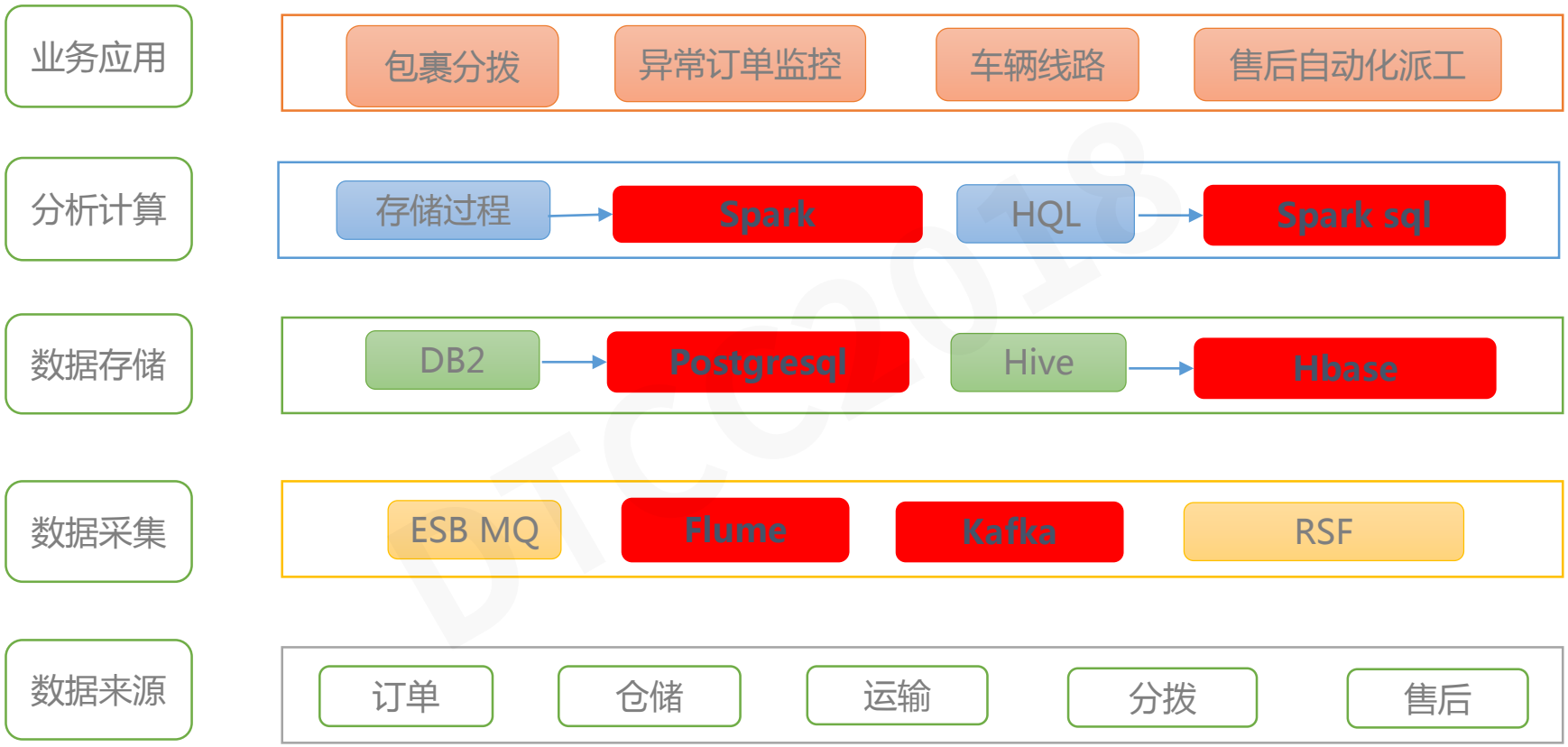
架构1.0



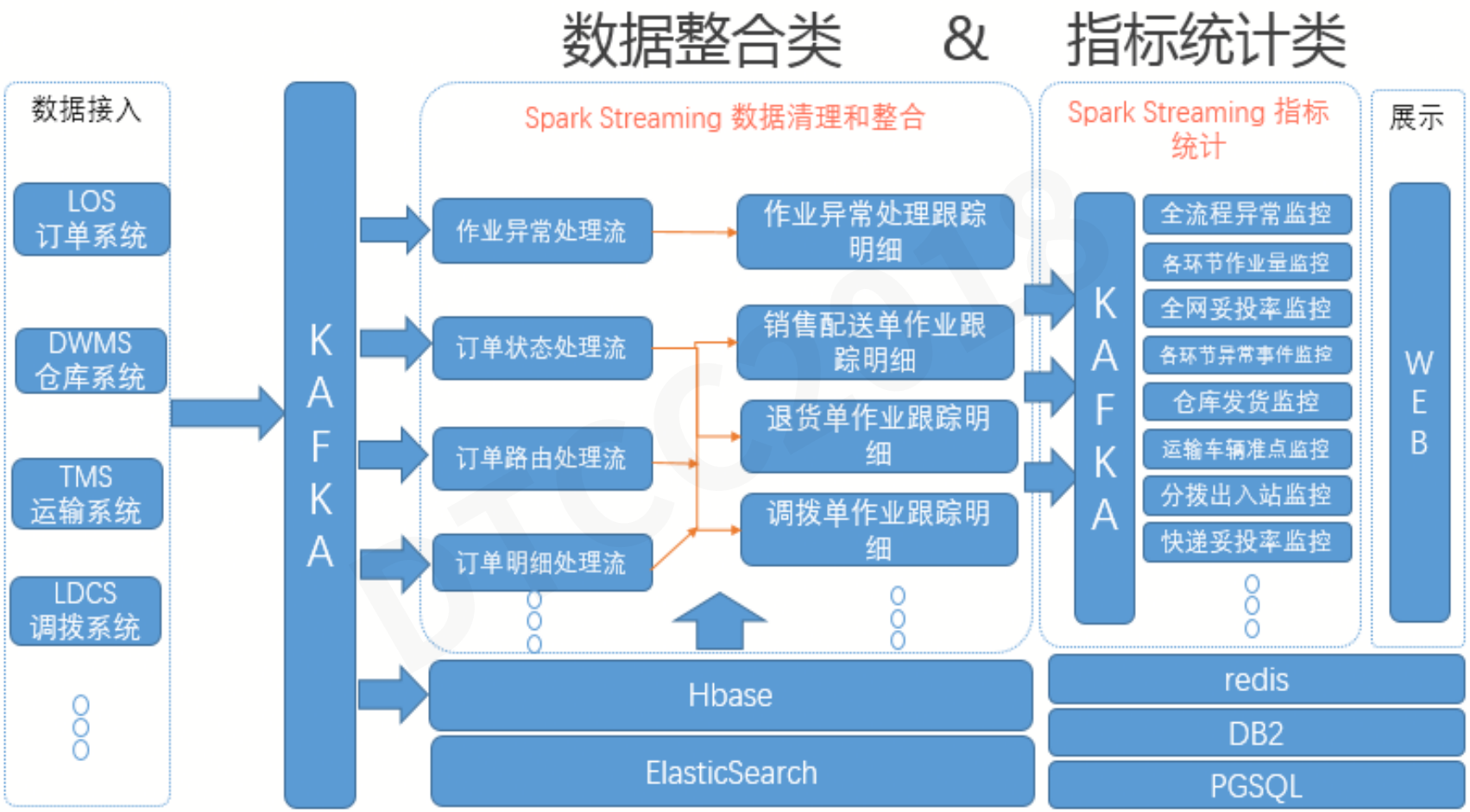
架构2.0计算选型

Spark	Storm
• 吞吐量高	• 吞吐量低
• 秒级计算(流)	• 毫秒级(逐条)
• 扩展方便	• 扩展方便
• java,sql	• clojure,java
• 高可用	• 高可用

架构2.0



架构3.0


数据整合类 & 指标统计类
数据接入
LOS 订单系统
DWMS 仓库系统
TMS 运输系统
LDCS 调拨系统
KAFKA
Spark Streaming 数据清理和整合
作业异常处理流
订单状态处理流
订单路由处理流
订单明细处理流
作业异常处理跟踪明细
销售配送单作业跟踪明细
退货单作业跟踪明细
调拨单作业跟踪明细
Hbase
ElasticSearch
Spark Streaming 指标统计
KAFKA
全流程异常监控
各环节作业量监控
全网妥投率监控
各环节异常事件监控
仓库发货监控
运输车辆准点监控
分拨出入站监控
快递妥投率监控
展示
WEB
redis
DB2
PGSQL

架构升级-效果展示

每天实时更新上百张实时报表，数据量百T左右

实时性能从小时级别提高到了5分钟以内

1分钟处理的数据量达到上千万

目录

数据存储方案

Citus+Postgresql

Elaticsearch+HBase

数据库存储负载

更新

- 每5分钟更新约400张明细表
- 最宽的表600字段,5KB/行
- 最宽的表每次更新约400w记录

计算

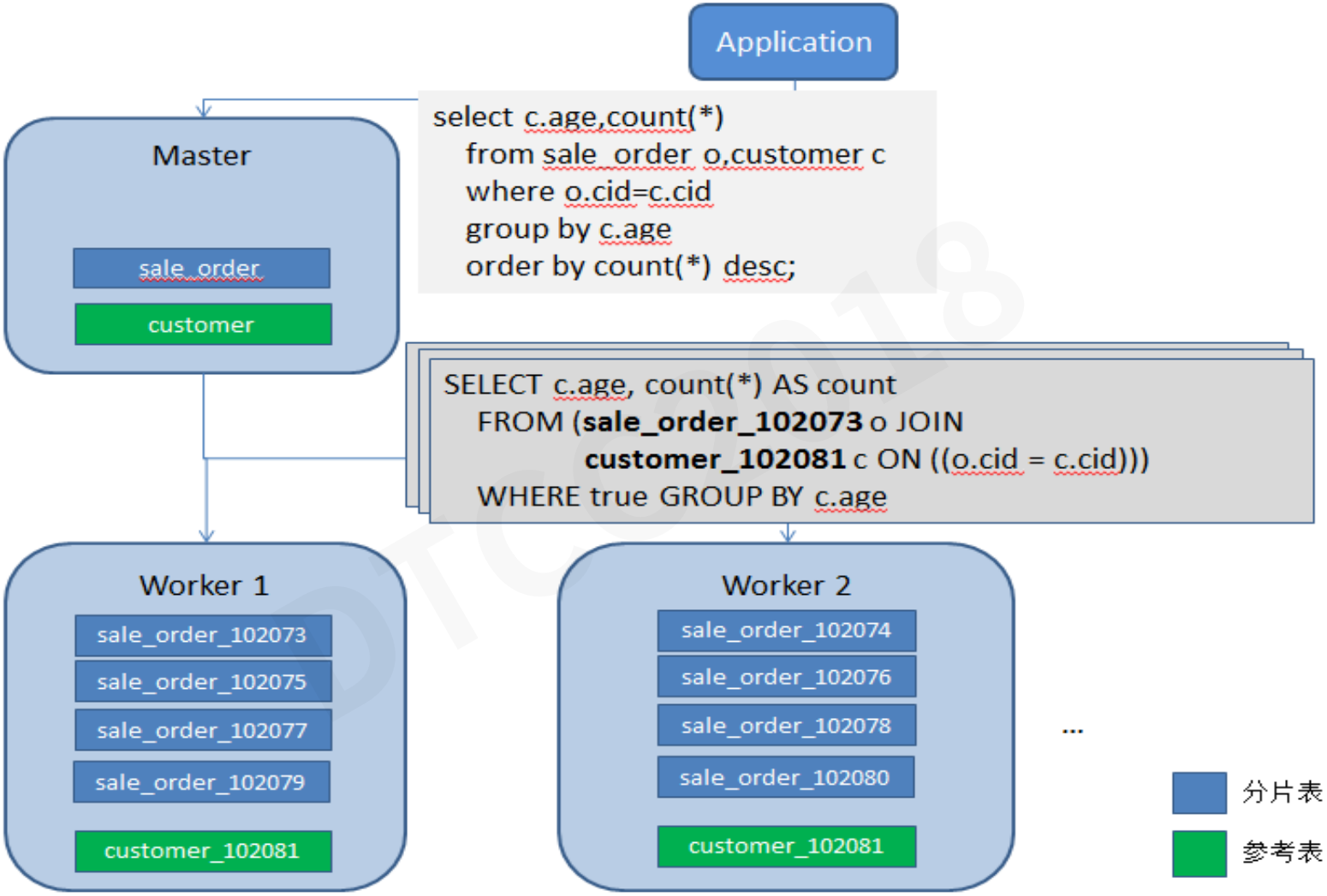
- 每5分计算上千张报表(2分内完成)
- 4000+次/min明细表聚合汇总运算

查询

- 要求并发度>400
- 实时性要求在3秒内

十倍增速

Citus


Application
select c.age,count(*)
from sale_order o,customer c
where o.cid=c.cid
group by c.age
order by count(*) desc;
Master
sale_order
customer
SELECT c.age, count(*) AS count
FROM (sale_order_102073 o JOIN
customer_102081 c ON ((o.cid = c.cid)))
WHERE true GROUP BY c.age
Worker 1
sale_order_102073
sale_order_102075
sale_order_102077
sale_order_102079
customer_102081
Worker 2
sale_order_102074
sale_order_102076
sale_order_102078
sale_order_102080
customer_102081
...
分片表
参考表

数据存储方案-Postgresql

Master元信息

pg_dist_partition

- 存储数据库表哪些是分片表

pg_dist_shard

- 存储分片表有哪些分片组成

pg_dist_shard_placement

- 存储的分片实际在哪台机器上

CitusDB执行器

Real-time

- Master与后端所有shard建立连接
- 快速响应，实时性强

Task Tracker

- Master只与worker建立一个连接
- 实时性差

Routing

- 处理Insert、update、delete操作

SQL限制

Join限制

- 不支持2个非亲和分片表的outer join
- 仅task-tracker执行器支持2个非亲和分片表的inner join
- 对分片表和参考表外连接，参考表只能出现在left join的右边或right join的左边

子查询、插入限制

- 子查询不能参与join
- 子查询不能出现limit、offset
- 插入分片表，master负载过重，插入效率低

解决方案

- 通过临时表(或dblink)中转
- 查询记录所属分片位置，直接与worker建立连接进行插入

明细下载

- 并发下载量1万+
- 全国分拨中心和10万+快递点全覆盖
- 最大下载数据量百万+
- 实时下载时长控制在30秒内

明细下载策略

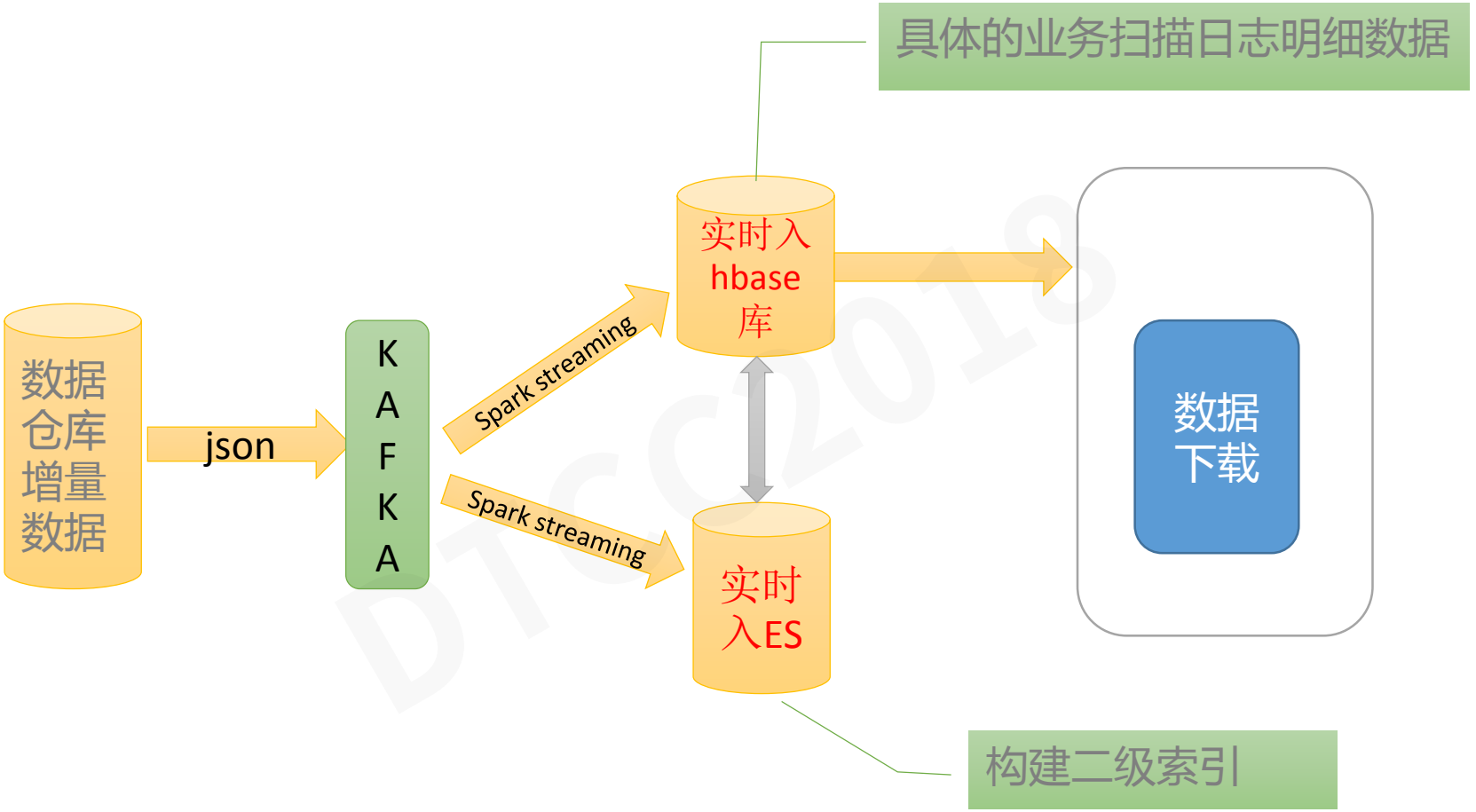
实时明细下载(小数据量)

- Postgresql

历史明细下载(大数据量)

- Postgresql同步数据到Hbase+ES
- Hbase存储数据明细
- Elaticsearch关键条件字段索引

数据下载技术架构



目录

经验分享

- 数据倾斜案例分享
- Spark调优总结
- PostgreSQL经验分享

数据倾斜-场景一

场景

- 大量数据集中在某几个Key上

方案

- 直接剔除掉

优缺点

- 方法简单，完全规避了数据倾斜
- 适用场景不多

数据倾斜-场景二

场景

- 对RDD执行reduceByKey等聚合算子或者在Spark SQL中使用group by语句进行分组聚合时

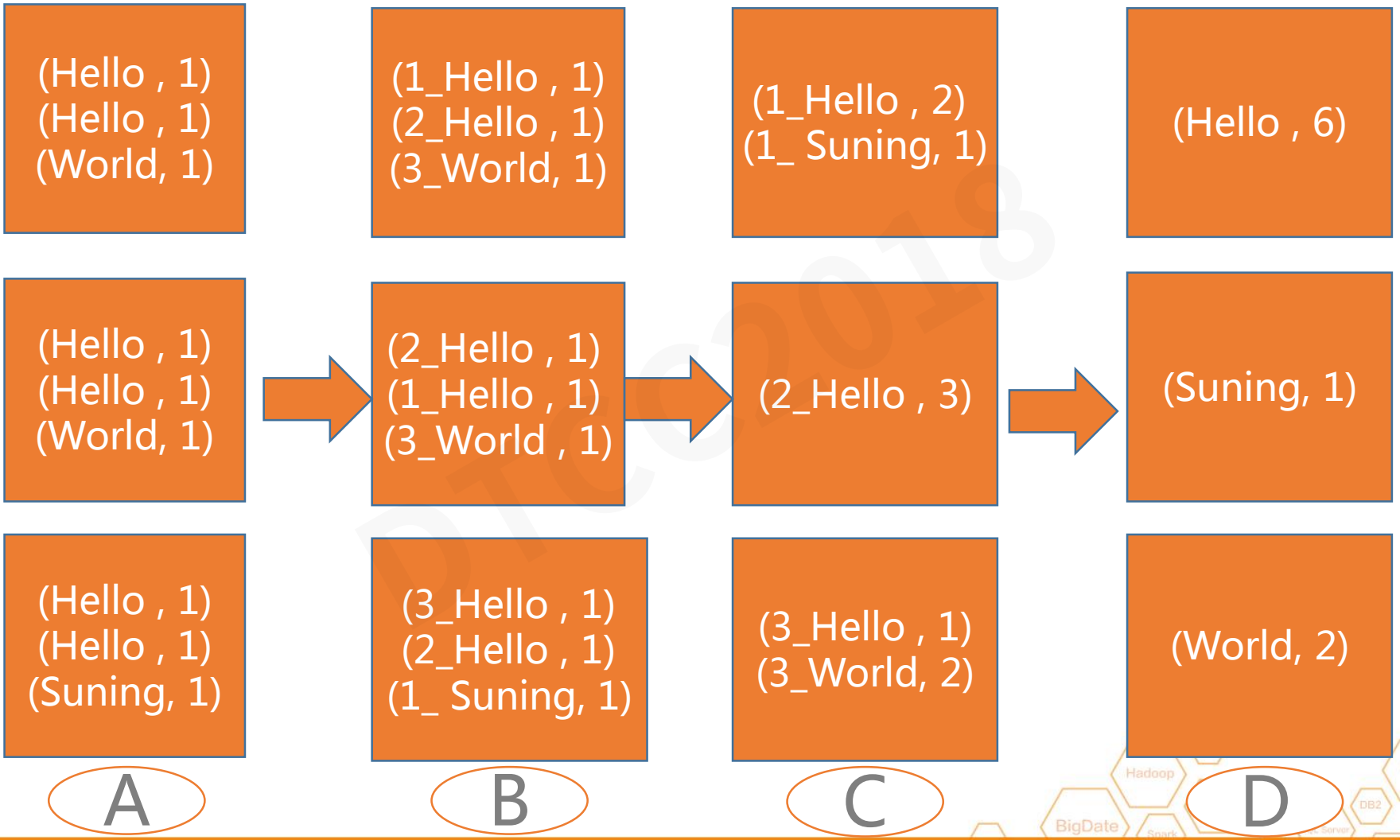
方案

- 两阶段聚合（局部聚合+全局聚合）

优缺点

- 通常都可以解决掉聚合类数据倾斜
- 仅适用于聚合类的shuffle操作

数据倾斜-场景二



数据倾斜-场景三

场景

- 在表数据使用join类操作时，其中一个表的数据量比较小，另外一个数据量较大

方案

- 使用Broadcast变量与map类算子代替join

优缺点

- 优点：对join操作导致的数据倾斜，效果好
- 缺点：只适用于一个大表和一个小表的情况

数据倾斜-场景四

场景

- 在表数据使用join类操作时，其中一个表的大量key导致数据倾斜，另外一个表数据分布正常

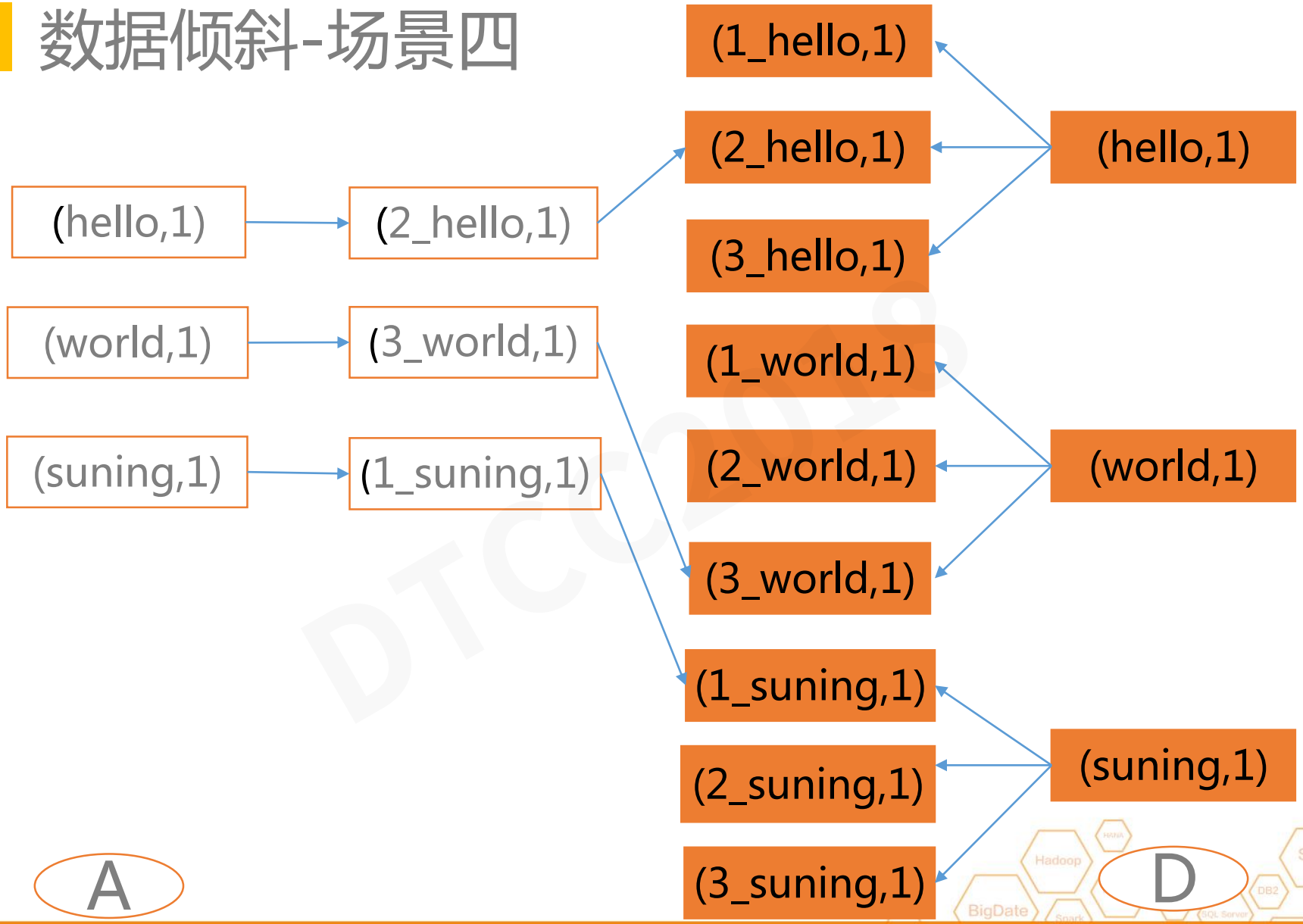
方案

- 对倾斜表添加随机前缀
- 扩容数据分布正常表

优缺点

- 优点：对join类型的数据倾斜基本都可以处理，而且效果也相对比较显著
- 缺点：该方案更多的是缓解数据倾斜，而不是彻底避免数据倾斜，扩容表需要更多内存

数据倾斜-场景四



Spark开发调优

1. 尽量避免或尽量少的使用shuffle算子

易购

2. 对多次使用的RDD或DataFrame进行缓存，共享同一个RDD

3. 分区调整
- 经过filter算子过后使用coalesce优化分区数量。
- 分区少并且数据量大是通过repartition重分区增大并发。

4. 使用foreachPartition代替foreach,使用mapPartition代替map。

5.使用spark.streaming.kafka.maxRatePerPartition限流

Postgresql调优

1. where 条件尽量少用函数
   - coalesce(c,’ ’) <>’ x’ 改成 c is null or c<>’ x’
   - coalesce(c,’ ’)= ‘’ 改成 c is null or c= ‘’
   - coalesce(c,’ ’)<> ‘’ 改成 c is not null and c<> ‘’
   - substr(c,1) = ‘L’ 改成 c like ‘L%’

2. master(real-time)到worker用的短连接，pgbouncer(数据连接池)默认记录连接和断连接事件，导致日志文件增长太快。后将其关闭

3. 插入数据采用跳过master直接插入worker 的做法

苏宁物流-未来展望

THANKS


HANA
Hadoop
SQL
MySQL
DB2
BigDate
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