

# 挑战双11实时数据洪峰的流计算实践

阿里巴巴数据技术及产品部 陈同杰



1

阿里流计算介绍

2

面临的挑战

3

我们是如何做的？

# 阿里巴巴流计算介绍





## 双十一媒体直播大屏





商家统一数据平台：生意参谋

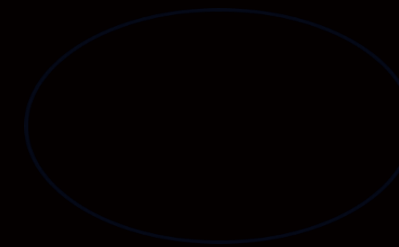
## 阿里数据的现状



100M Events/s  
每秒一亿记录



100B Events/day  
一天万亿记录



PB  
Everyday



EB  
Total

# 面 临 的 挑 战

## 技术难点



Low Latency  
低延时



Exactly-Once  
高精度



High Throughput  
高吞吐



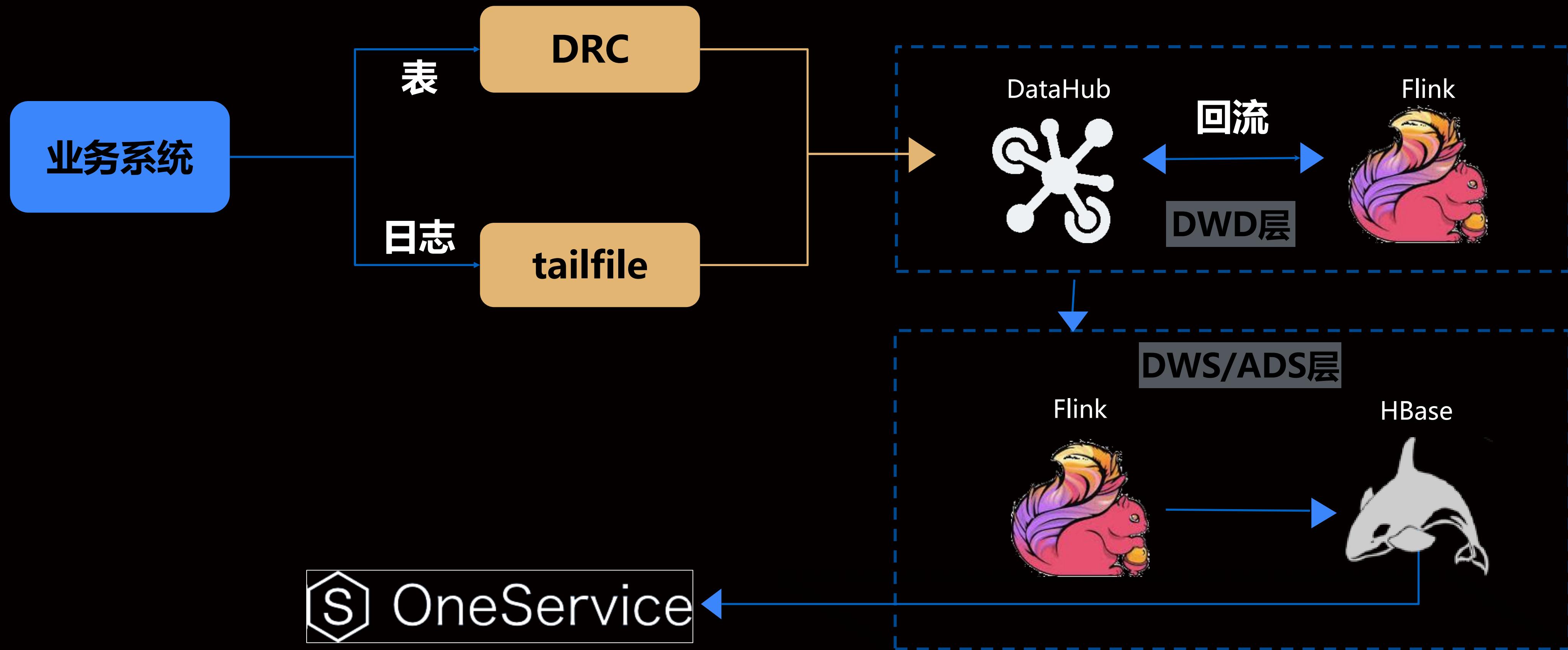
Strict SLA  
强保障



# 阿里巴巴如何做流计算



## 数据链路





## 流 计 算 引 擎 对 比





## What is Blink ?



Flink

+



Alibaba Group

=

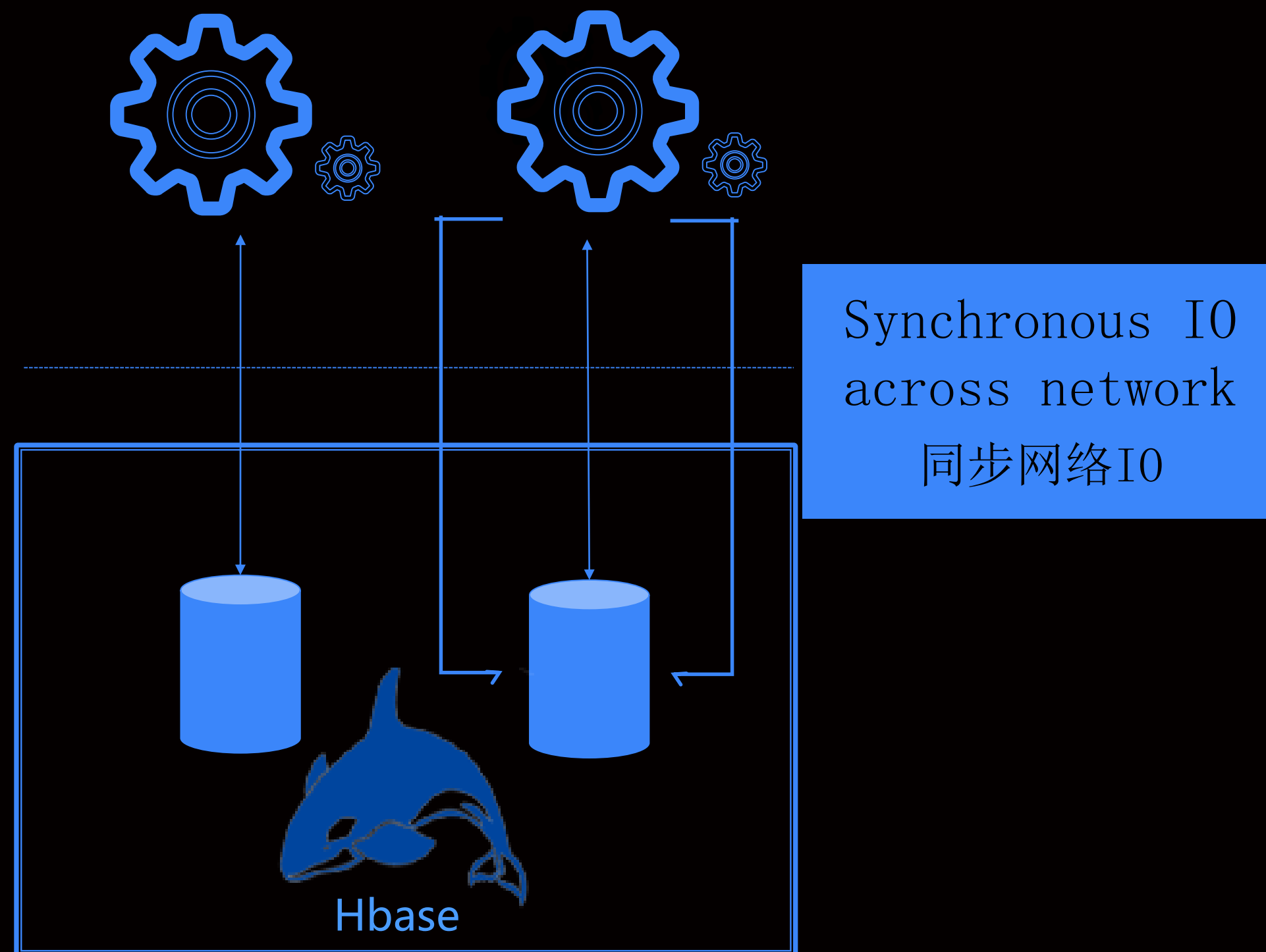


Blink

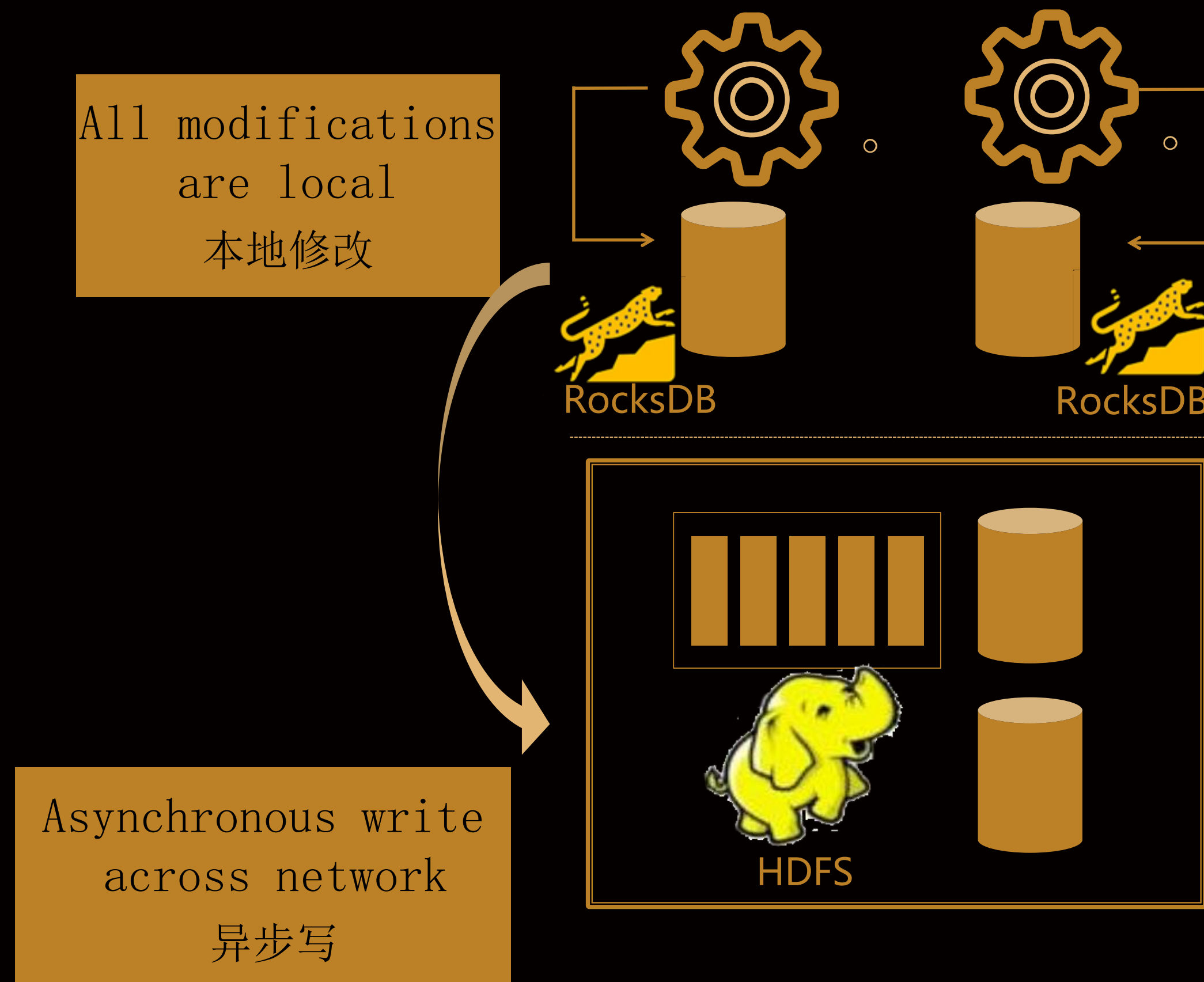


# Why Blink ?-Stateful Processing

Classic Architecture

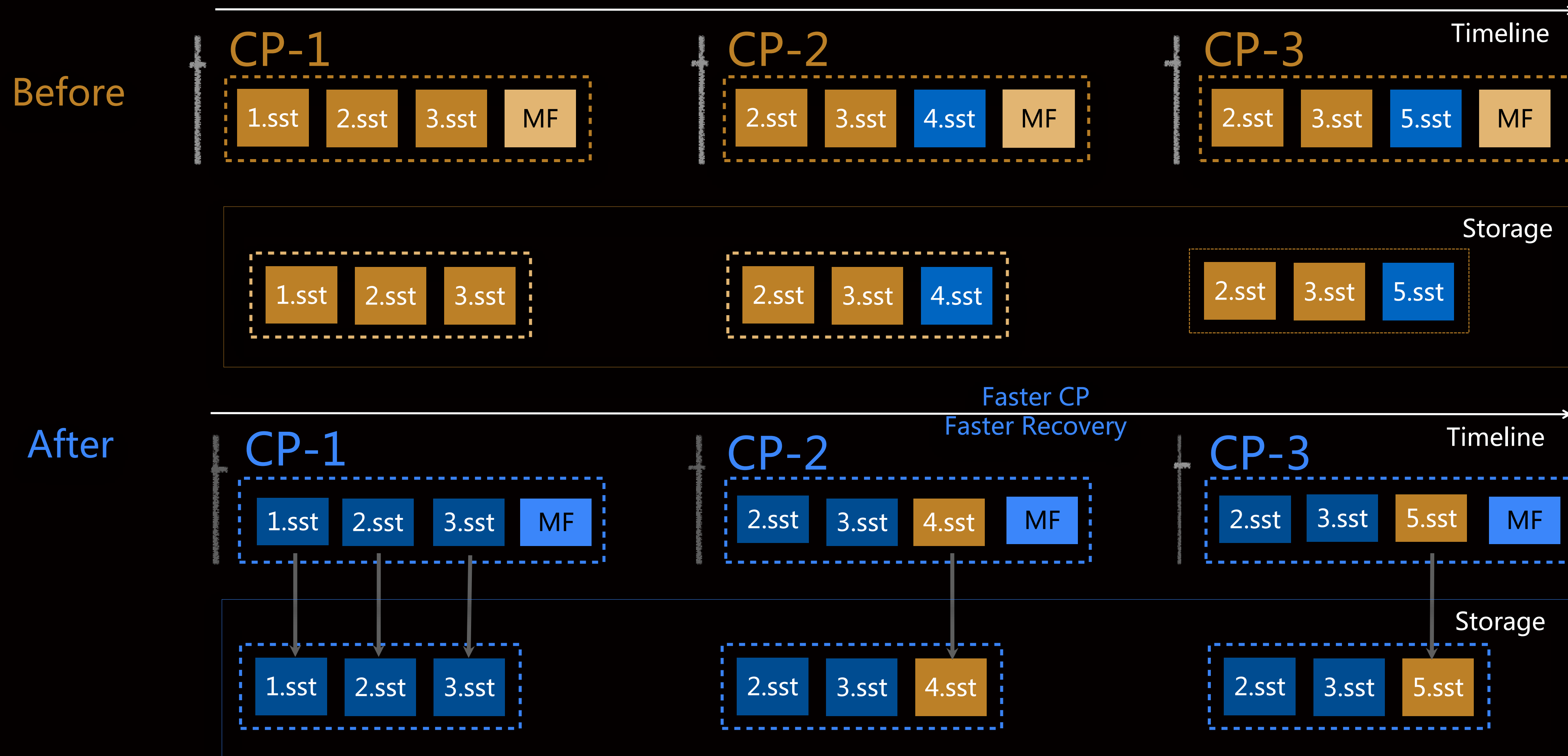


Blink with rocksDBStateBackend



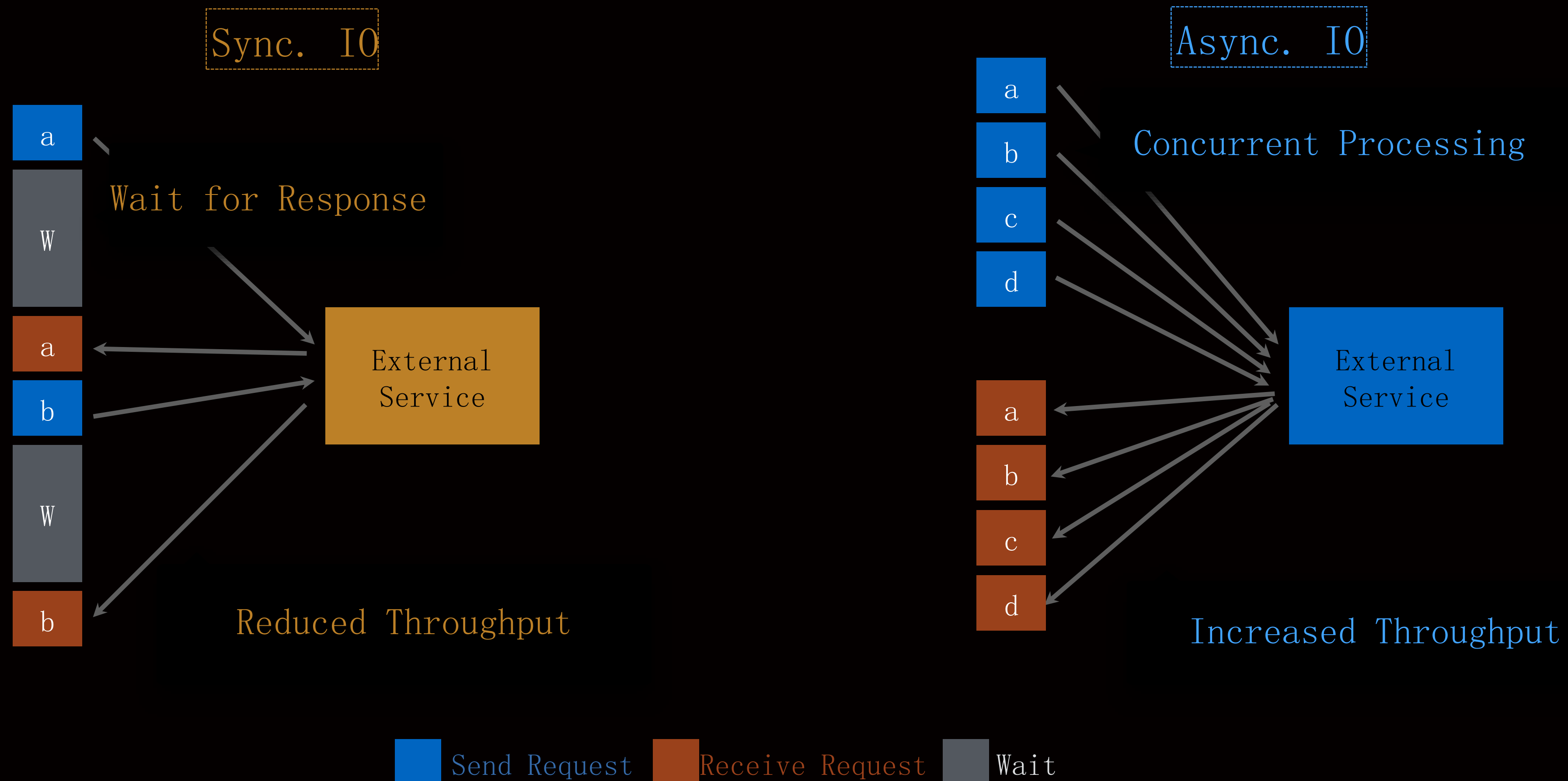


## Why Blink?-Incremental Checkpoint



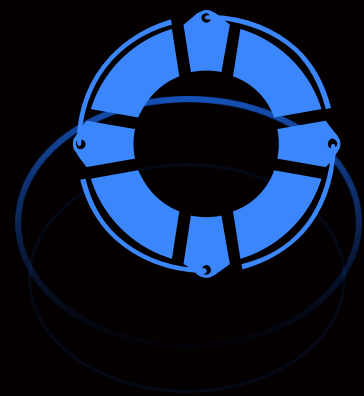


# Why Blink?-Asynchronous IO





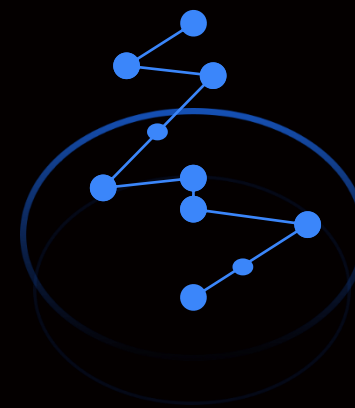
And Many More...



纯流式引擎



Checkpoint机制



流控与反压



实时监控



大规模部署



## 聚合组件

1

### 大小维度合并

减少网络传输50%以上

2

### 精简存储

利用index来存储指标，state存储减少一半

3

### 高性能排序

top组件利用PriorityQueue + MapState，大幅减少序列化次数，性能提高10倍左右

4

### 批量写操作

mini-batch sink，降低 HBase 压力

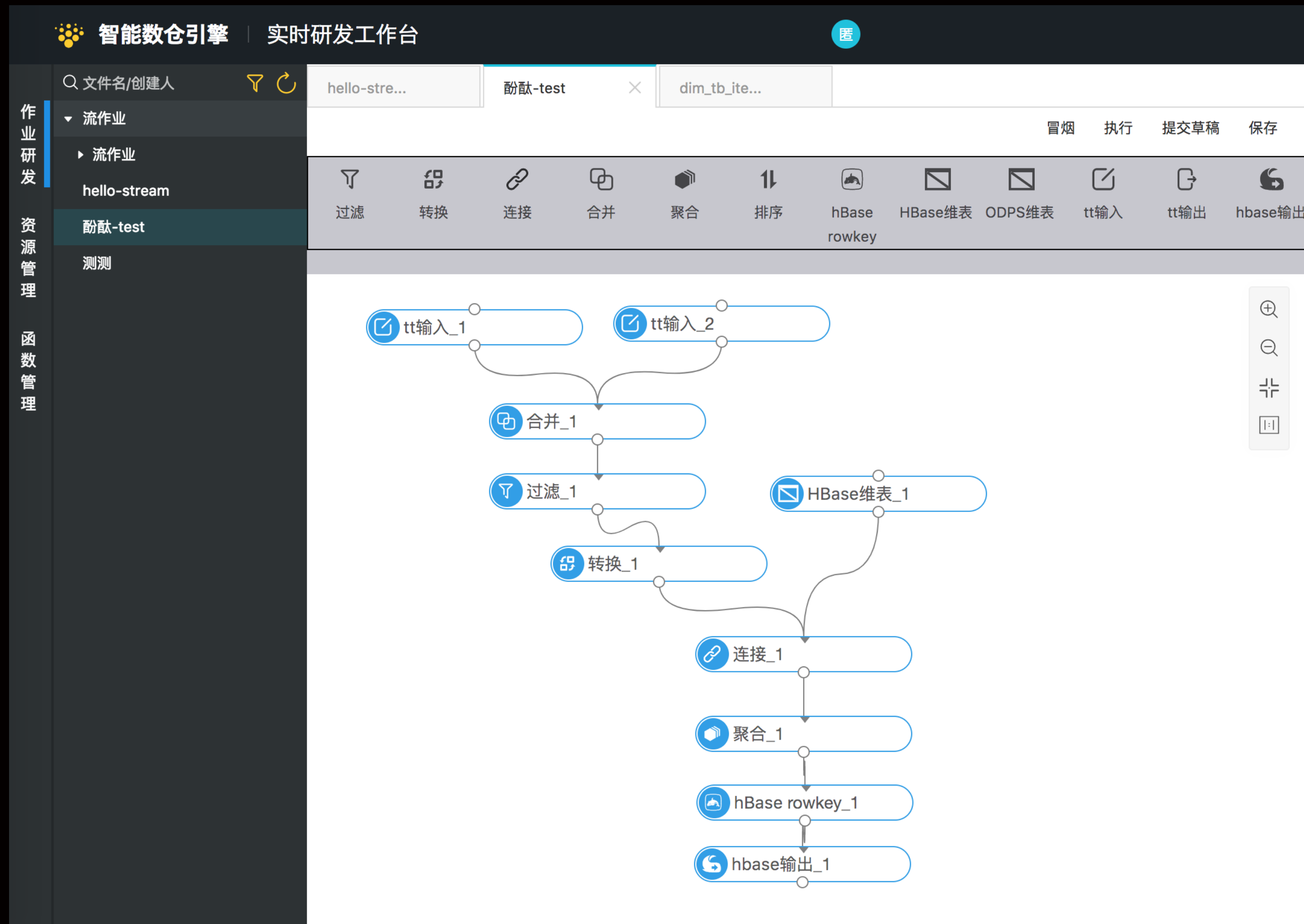
5

### 多条件分支优化

大幅减少网络传输与state大小



# 流计算开发平台 —— 赤兔



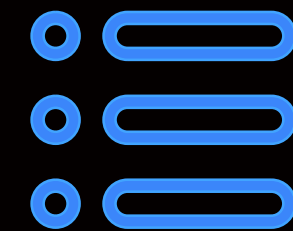
# Beam , TableAPI and SQL



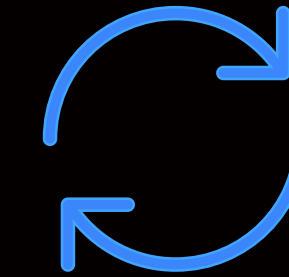
## Streaming Into Future ...



**Portal, Stream  
Processing as a Service**  
服务化平台



**Beam  
TableAPI & SQL**  
语义层统一



**Stream & Batch  
Unification**  
实时离线统一



**Machine Learning  
in real time**  
实时智能



# THANK YOU



扫码立即了解更多“阿里数据”精彩内容



# 阿里大数据专场

Alibaba Big Data Session

赋能商业，创造价值

Boost Business, Build Value

指导单位  
浙江省人民政府

主办单位  
杭州市人民政府  
阿里巴巴集团  
蚂蚁金服集团

承办单位  
杭州市西湖区人民政府  
杭州市经济和信息化委员会

杭州市西湖博览会组织委员会办公室  
阿里云计算有限公司  
云栖联盟

战略合作伙伴

