

# 从JStorm到Flink：腾讯实时流计算平台的建设与实践



# 目录

**01**

腾讯实时计算概况

**02**

Oceanus平台介绍

**03**

开源特性增强

**04**

流计算业务实践



# 实时计算业务赋能





# 腾讯实时计算规模



**2.1亿/秒**

峰值消息处理



**20万亿**

日均消息总条数



**3PB**

日均消息总大小

# JStorm到Flink的演进历程

2017.03

Flink框架预研, JStorm上K8S

新框架调研, 评估Flink替代JStorm的可行性, 包括: 功能、性能等的对比。此时所有的流计算任务全部由JStorm承载, 且继续演进新版本。

2018.01

Flink产品化

开始建设基于Flink的一站式流计算平台Oceanus, 并针对性优化社区Flink On Yarn模式的部分功能。

2019.03

Oceanus场景化建设

打造内部Oceanus生态, 完善场景化服务, 上线在线ML, 开始建设秒级监控等平台。

2017.09

Flink开源版本增强, 业务迁移

针对内部场景, 部分改造优化开源版本, 开始迁移部分JStorm上的任务, 以standalone模式运行

2018.09

Oceanus平台上线

实时流计算平台规模化接入腾讯内部业务(覆盖所有BG)、TBDS客户, 上线公有云。开始批量迁移存量JStorm的任务到Oceanus。内部宣布JStorm版本不再演进。



# 目录

**01**

腾讯实时计算概况

**02**

Oceanus平台介绍

**03**

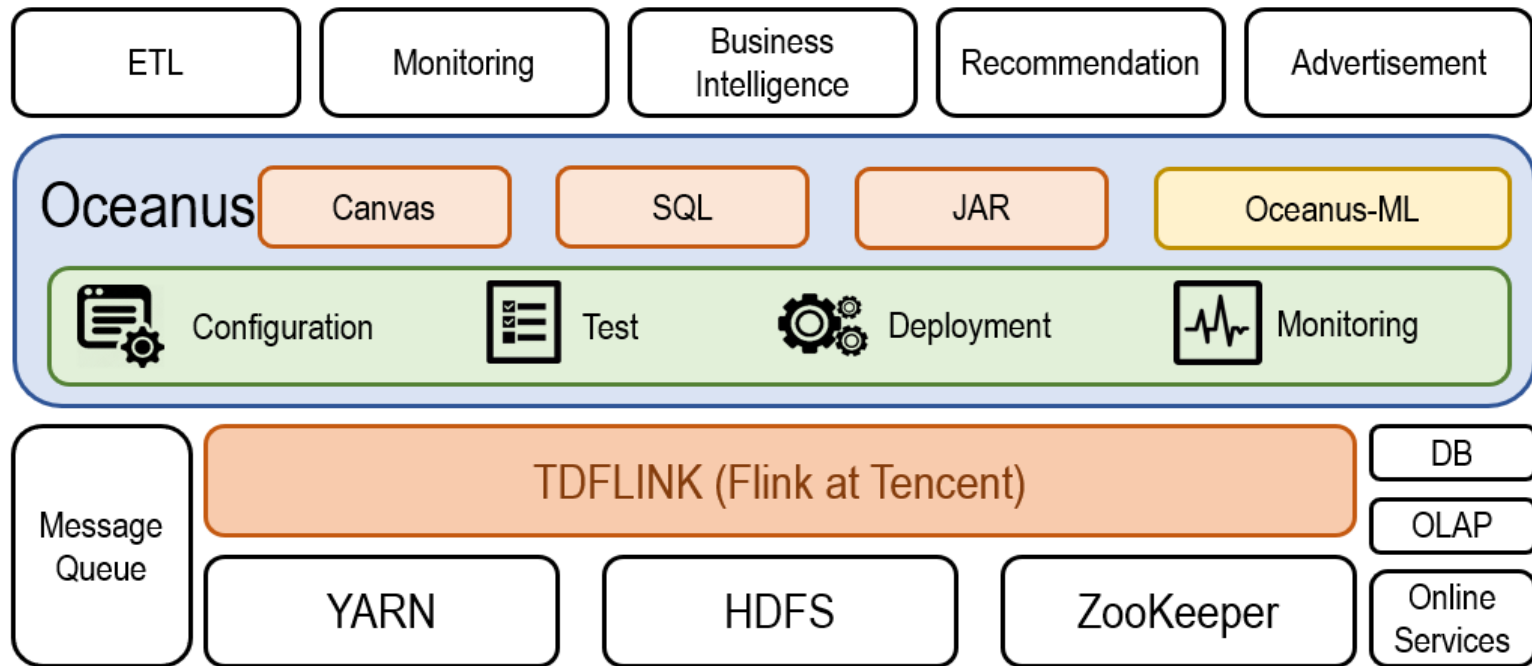
开源特性增强

**04**

流计算业务实践



# Oceanus实时计算平台



# 平台建设重点



## 一站式

提供完善的上下游数据生态，形成从接入、处理，到应用的数据闭环。



## 自助化

借助平台的应用调试、丰富的实时指标数据、配置化告警等功能，实现监控、运维自助化。



## 可视化

不仅通过平台Web页面可轻松查看指标、打点日志等数据，同时对调试结果，应用输出等，都提供了可视化的方式。



## 持续增强

结合业务实践，持续优化既有能力，并且迭代新功能；在满足自身客户需求的同时，适时回馈社区。

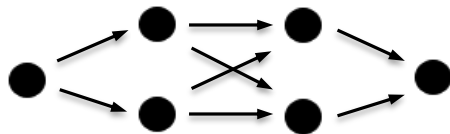




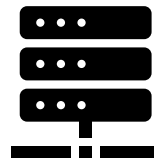
# Oceanus-提交计算任务



配置元数据




创建DAG



编译提交



# Oceanus-在线调试



🔍 📄 ⓘ 📧 ⓘ

应用管理 > 调试

1 配置调试数据

2 调试结果

应用信息

应用名:  [查看SQL](#)

责任人:

源表数量: 2

调试数据 ⓘ

Source源表	操作
<div><div>+</div><div>t_ec00</div></div>	<a href="#">随机生成</a> <a href="#">上传数据</a>
<div><div>+</div><div>v_dc</div></div>	<a href="#">随机生成</a> <a href="#">上传数据</a>

<

1

>

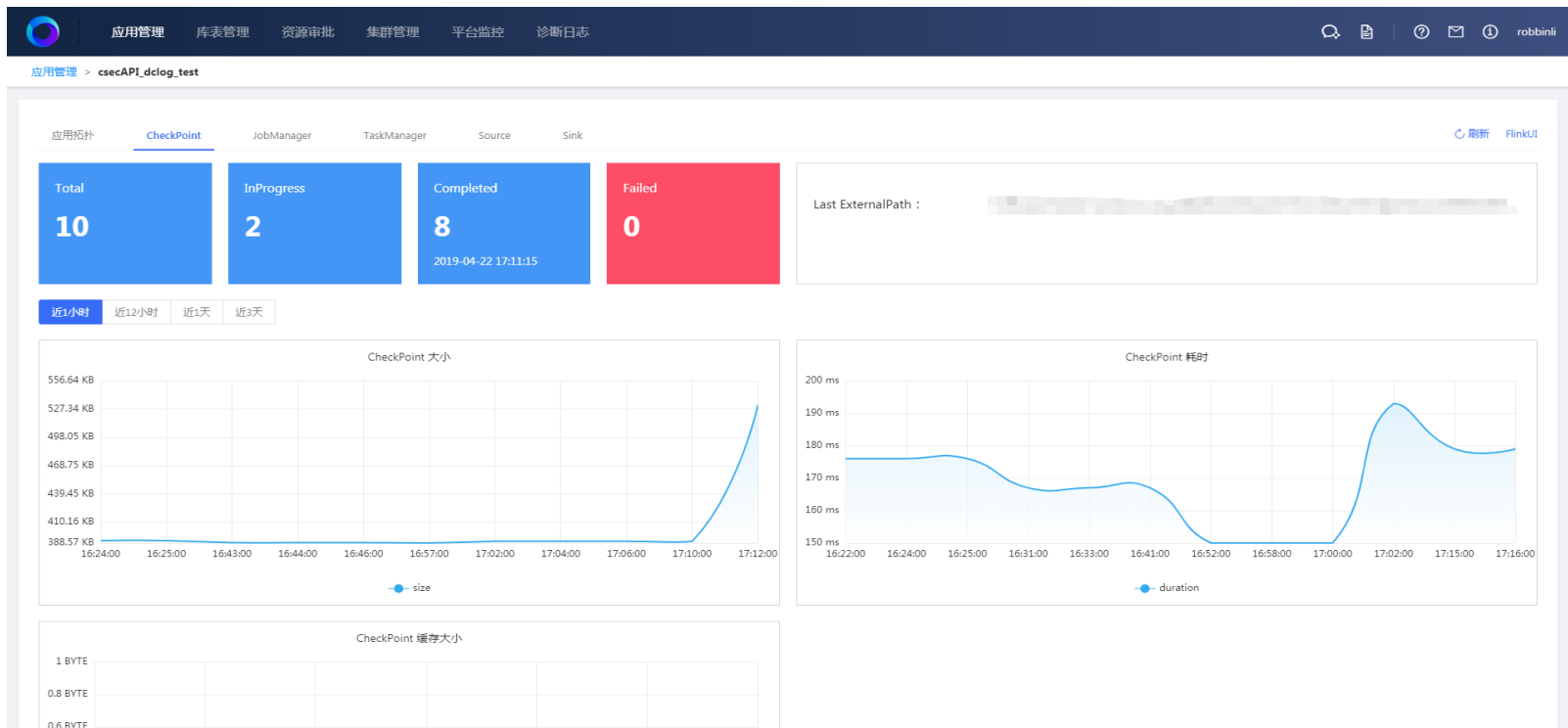
取消

开始调试

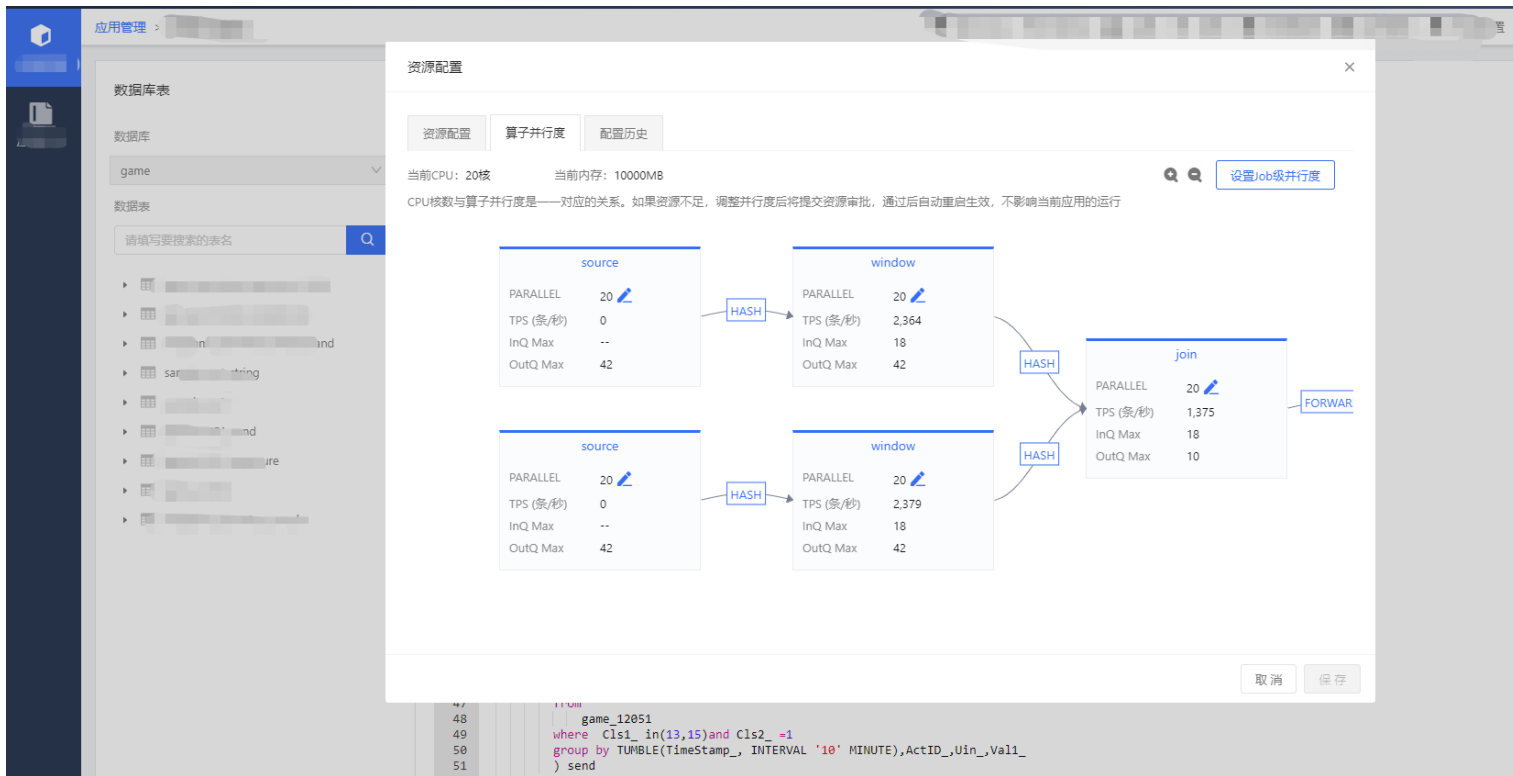
Copyright © 1998 - 2019 Tencent. All Rights Reserved. 腾讯公司 数据平台部 版权所有



# Oceanus-指标统计



# Oceanus-并行度调整



The screenshot displays the Oceanus web interface with a modal window titled "资源配置" (Resource Configuration). The modal has three tabs: "资源配置" (Resource Configuration), "算子并行度" (Operator Parallelism), and "配置历史" (Configuration History). The "算子并行度" tab is active, showing the configuration for a SQL job.

Current resources: 当前CPU: 20核 (Current CPU: 20 cores), 当前内存: 10000MB (Current memory: 10000MB).

Text: CPU核数与算子并行度是一一对应的关系。如果资源不足, 调整并行度后将提交资源审批, 通过后自动重启生效。不影响当前应用的运行 (The number of CPU cores and operator parallelism are one-to-one. If resources are insufficient, adjusting parallelism will submit a resource approval. After approval, it will automatically restart and take effect. It does not affect the current application's operation).

Buttons: 设置Job级并行度 (Set Job-level parallelism), 取消 (Cancel), 确定 (Confirm).

The diagram shows the execution plan with the following operator configurations:




Operator	PARALLEL	TPS (条/秒)	InQ Max	OutQ Max
source (top)	20	0	--	42
window (top)	20	2,364	18	42
source (bottom)	20	0	--	42
window (bottom)	20	2,379	18	42
join	20	1,375	18	10

The join operator is followed by a "FORWARD" label.

```
47  from
48      game_12051
49  where C1s1_in(13,15)and C1s2_ =1
50  group by TUMBLE(TimeStamp_, INTERVAL '10' MINUTE),ActID_,Uin_,Val1_
51  ) send
```




# Oceanus-自助诊断

 6 container_e07_15537	 35_01_000002	RUNNING	2019-07-11 21:17:52	2.6 day	2	0	2	1 GB
 56734 container_e07_4935_01_000003		RUNNING	2019-07-11 21:17:51	2.6 day	2	0	2	1 GB

< 1 2 3 >

5 条/页 跳至 页

概览 线程

734

ID	Name	State	CPU	Stack
8782	flink-metrics-23	TIMED_WAITING	0%	<a href="#">详情</a>
6570	flink-akka.actor.default-dispatcher-31	WAITING	0%	<a href="#">详情</a>
2540	flink-metrics-akka.remote.default-remote-dispatcher-20	TIMED_WAITING	0%	<a href="#">详情</a>
2470	flink-akka.remote.default-remote-dispatcher-30	WAITING	0%	<a href="#">详情</a>
2469	flink-akka.remote.default-remote-dispatcher-29	WAITING	0%	<a href="#">详情</a>

< 1 2 3 4 5 ... 44 >

5 条/页 跳至 页



# Oceanus-告警配置

The screenshot shows the Oceanus application management interface. On the left, there is a sidebar with '应用详情' (Application Details) and '应用日志' (Application Logs). The main area displays a list of databases and tables. A SQL query is visible in the background:

```
1 insert
2 into
3   game_
4 select
5   taskid as task,
6   cast(uin as bigint) as uid,
7   cast(touin as bigint) as tuid,
8   from
```

An '告警配置' (Alert Configuration) dialog box is open in the foreground. It contains the following information:

- Header: 告警配置
- Message: 告警消息将通过微信发给应用责任人
- Configuration items (告警配置项):
  - ☒ checkpoint失败
  - ☐ checkpoint大小阈值
  - ☐ checkpoint超时告警
  - ☒ Job自动重启告警
  - ☐ 正在运行Job数掉零
  - ☐ 可用网络缓存为零
  - ☐ 单个Source Task不消费数据告警 (仅当...)
  - ☐ 单个Sink Task不生产数据告警
  - ☐ 注册的TaskManager数量减少
  - ☐ 10分钟内TaskManager FullGC次数
  - ☐ 算子输入输出队列用满
  - ☐ 指标智能监控预警
- Buttons: 取消 (Cancel), 保存 (Save)



# 目录

**01**

腾讯实时计算概况

**02**

Oceanus平台介绍

**03**

开源特性增强

**04**

流计算业务实践

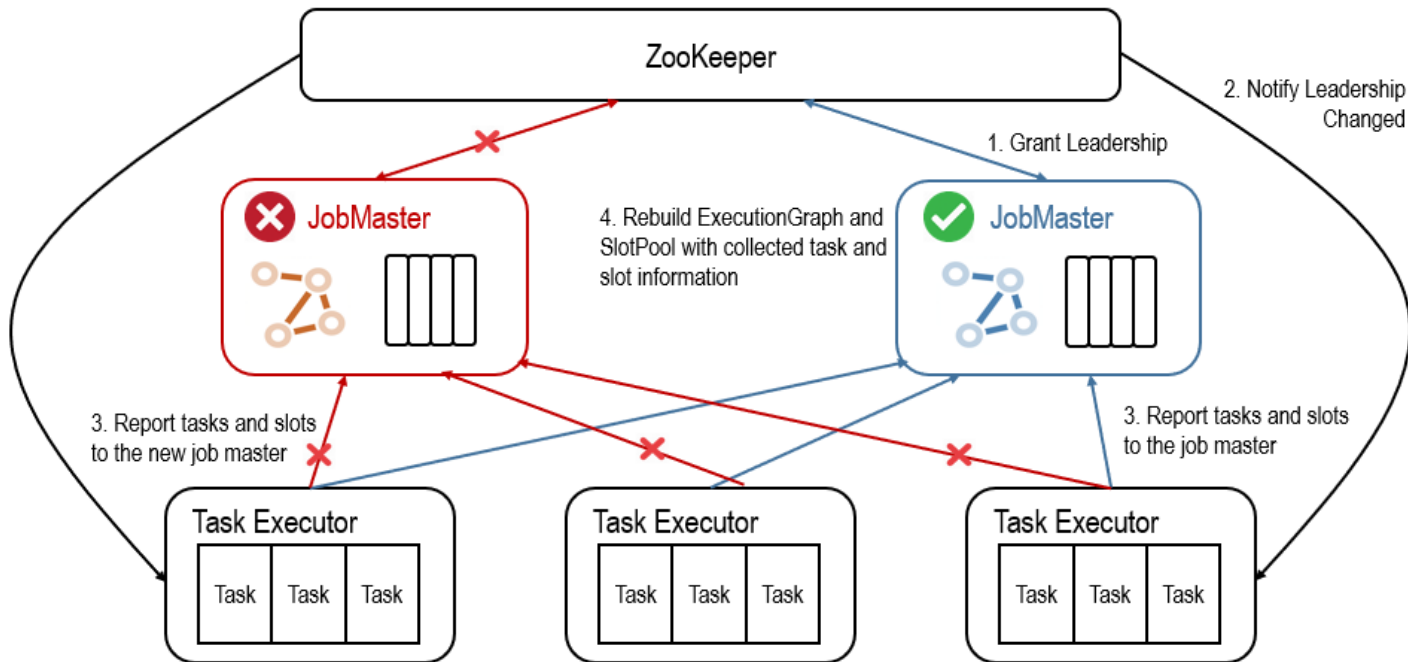


# 可靠性提升



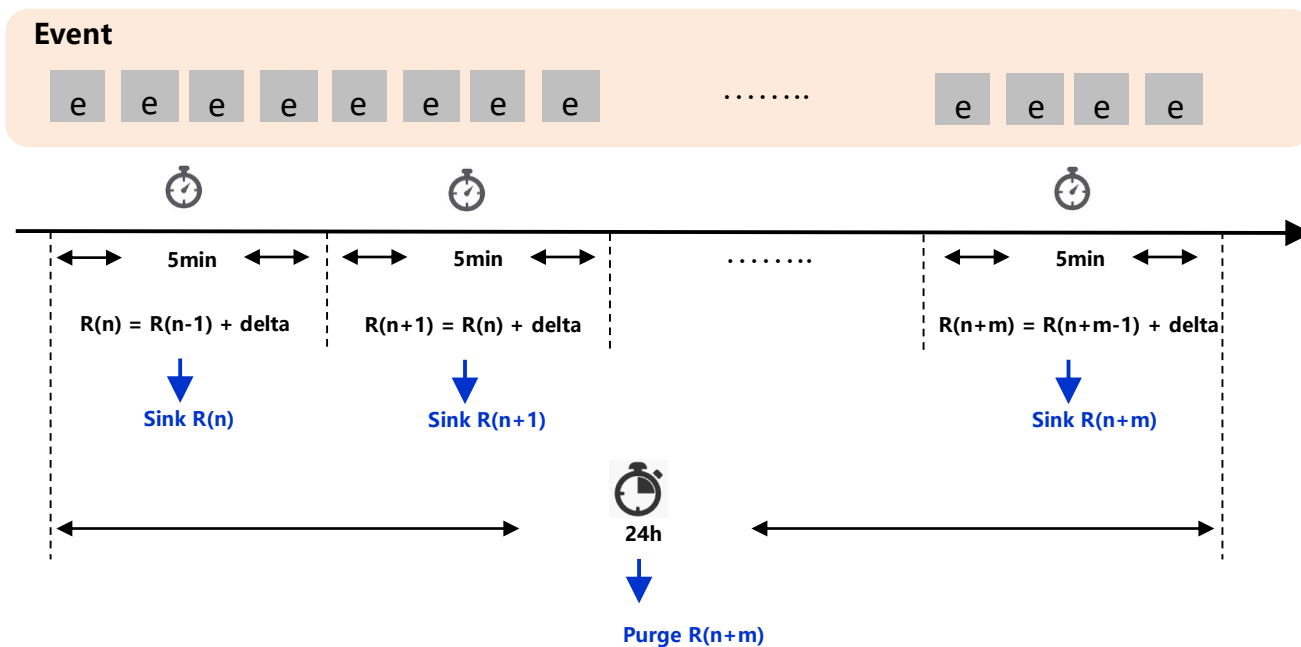
Recommend  
火龙果软件

Avoid restarting tasks when the job master fails.



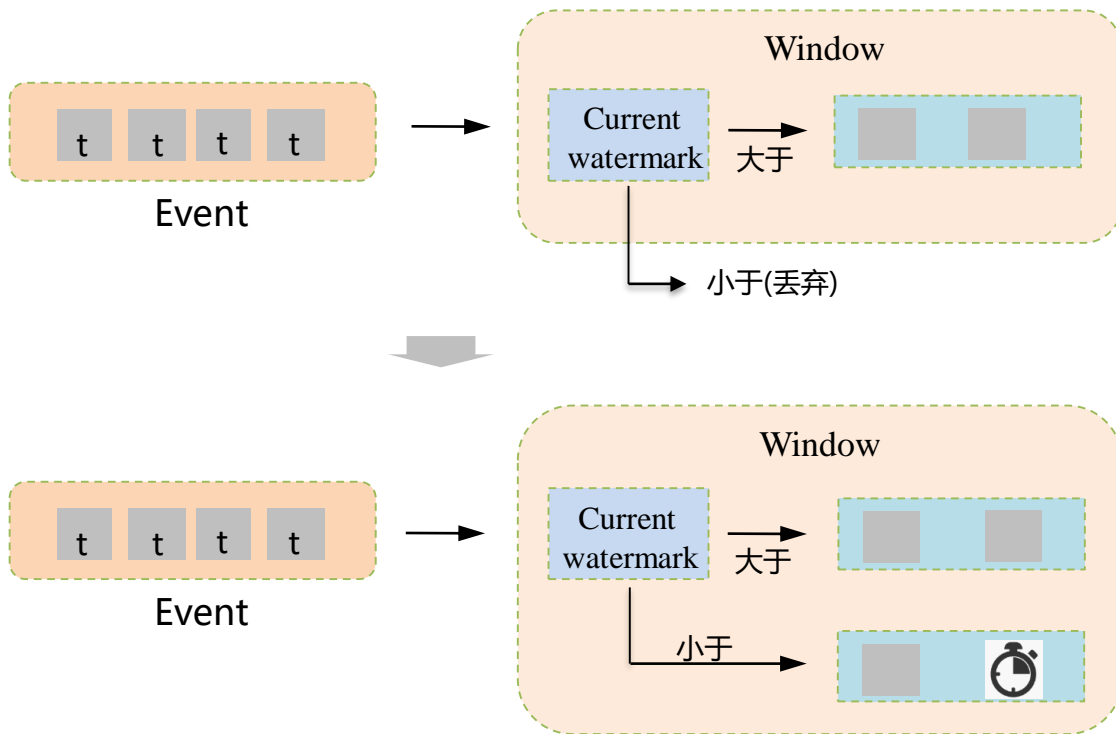


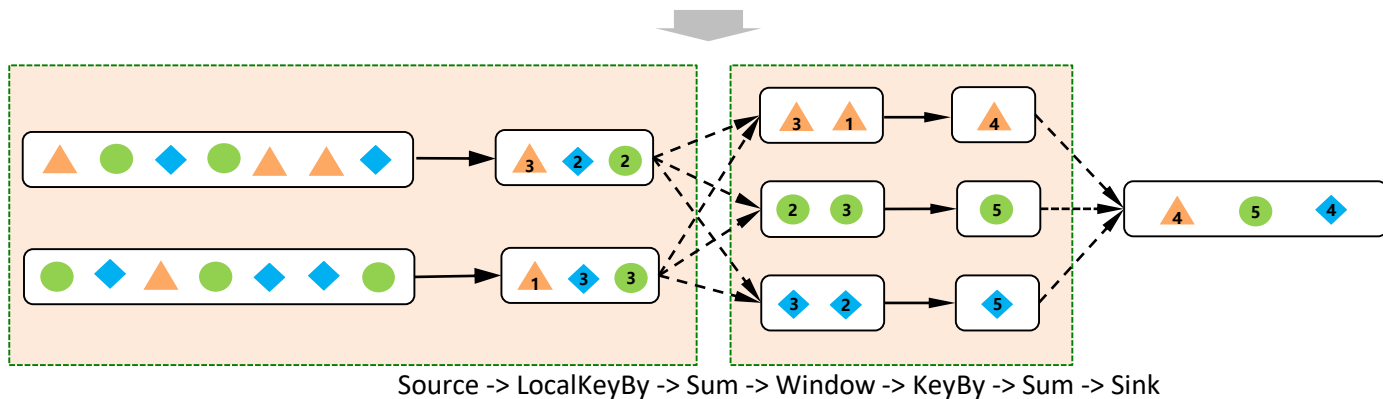
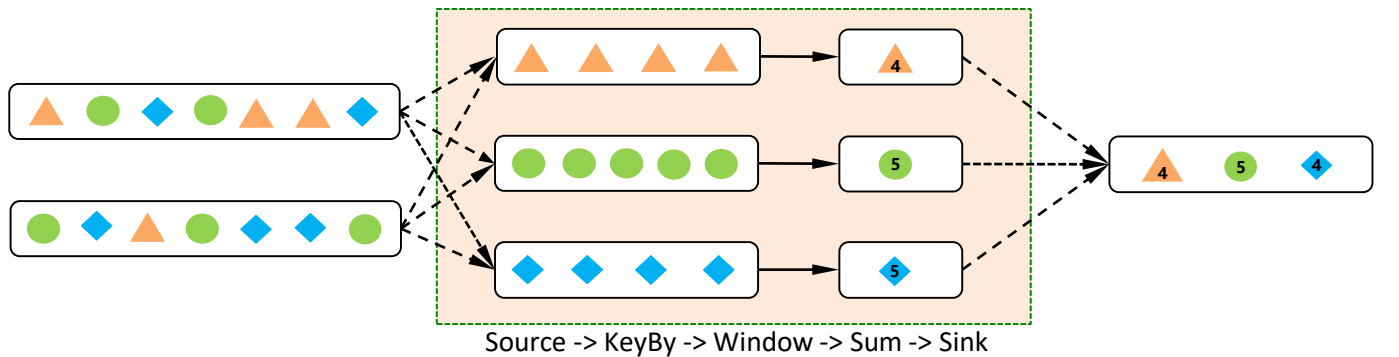
# Increment Window



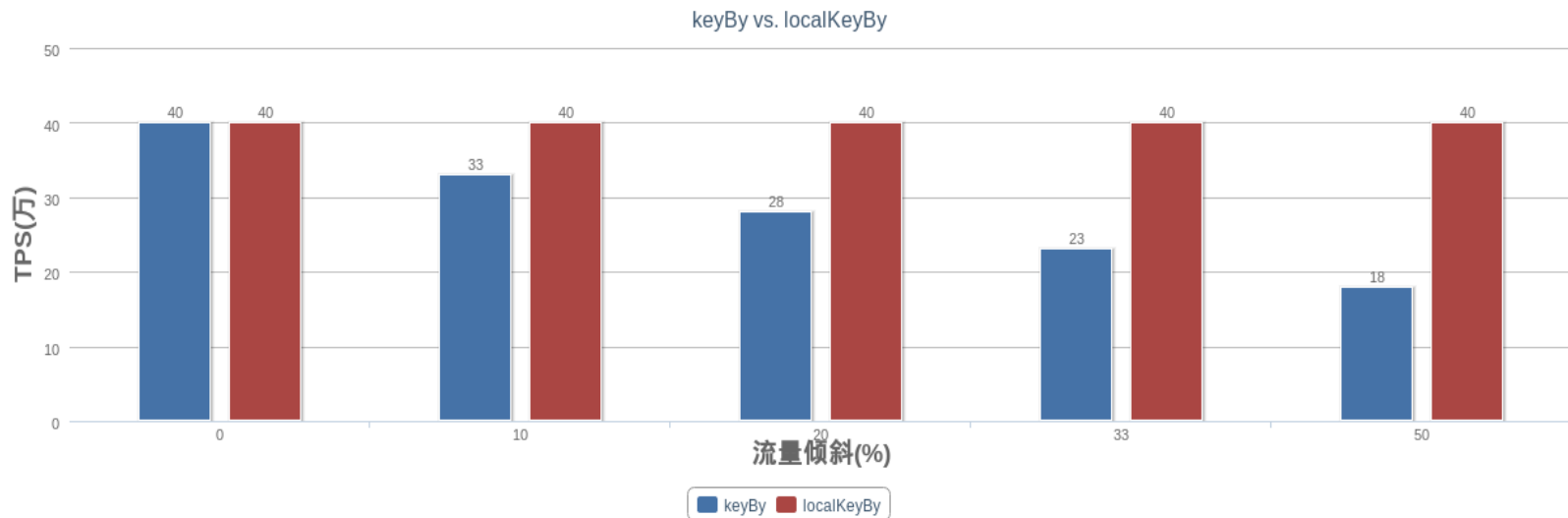


# Enhanced Window





# (Local)KeyBy对比





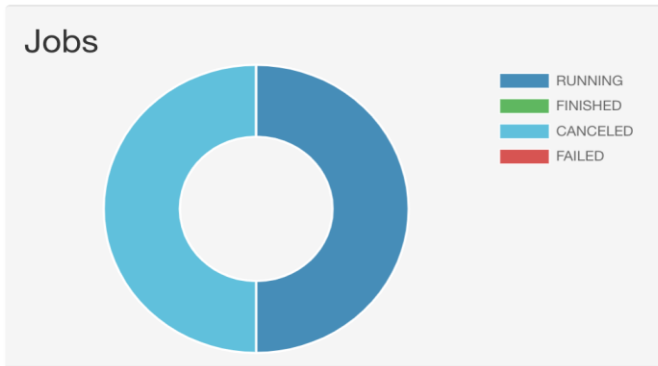
# Flink指标及UI重构

Task Managers  
1

CPU  
24 Cores

Memory  
62.6 GB

Slots (Available / Total)  
3 / 4



### Running Jobs

Name	Start Time	End Time	Duration	Parallelism	CPU	Memory	Status
<a href="#">Streaming WordCount</a> a9a0938e3dd64e25806e66851cbcdbad	2019-03-14, 9:12:47	N/A	15d 6h	2	0	0	<span>RUNNING</span>

### Completed Jobs

Name	Start Time	End Time	Duration	Parallelism	Status
------	------------	----------	----------	-------------	--------



# 目录

**01**

腾讯实时计算概况

**02**

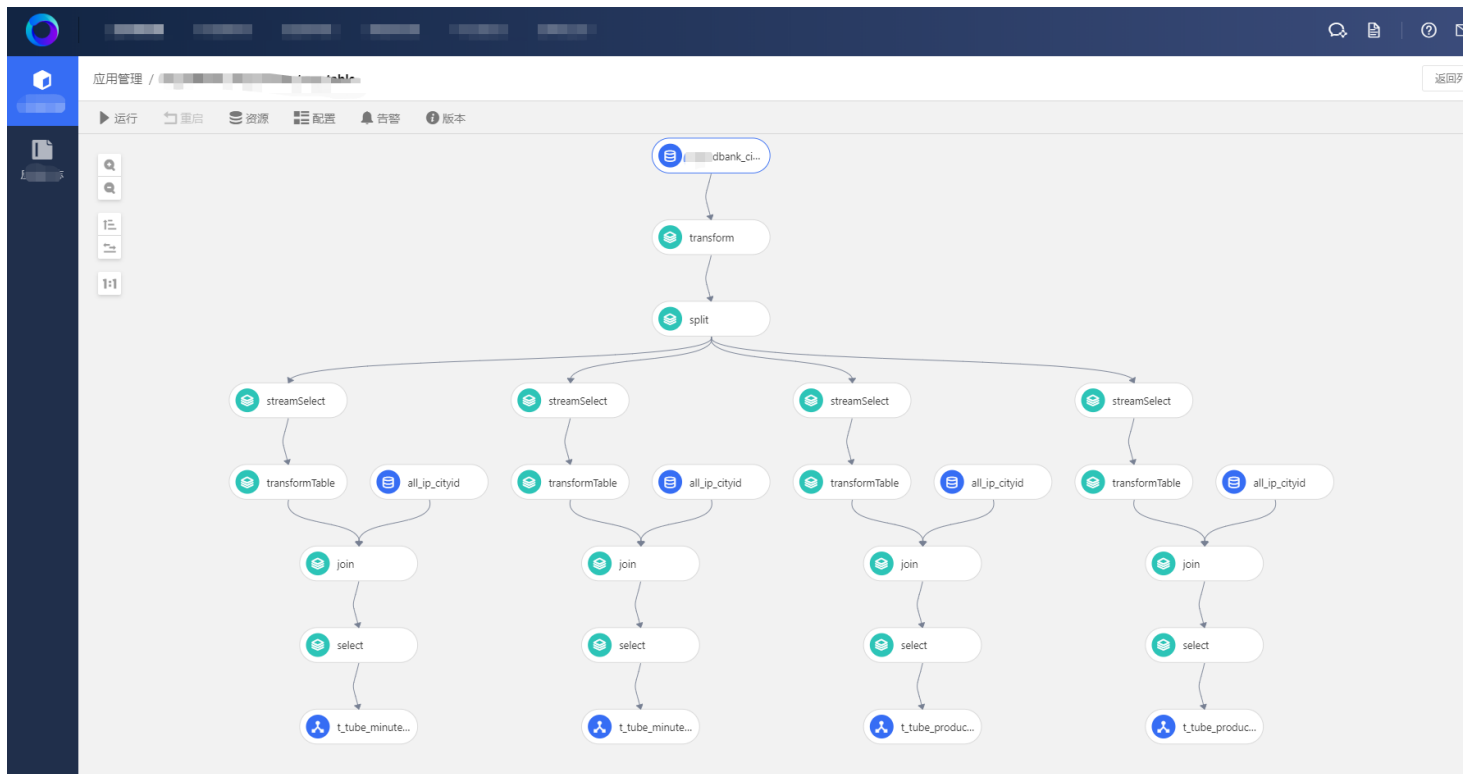
Oceanus平台介绍

**03**

开源特性增强

**04**

流计算业务实践



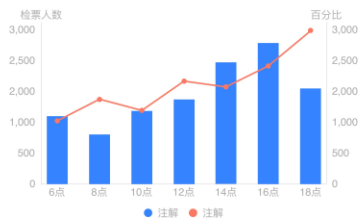


# 实时统计分析

腾讯小马BI可视化组件



腾讯小马BI可视化组件



腾讯小马BI可视化组件



腾讯小马BI可视化组件





应用管理

停止 重启 资源 配置 告警 版本

1-1

返回列表 编辑画布

CEP 配置

事件类型: Pattern (selected) Sequence

时间窗口: 90 秒

事件配置:

- 连续匹配: [enabled] 单一事件: [dropdown]  
e1 = [metricsValue > 0]
- 连续匹配: [disabled] 单一事件: [dropdown]  
e2 = [jobID == e1.jobID and metricsValue]

输出字段:

- e1.hostname as hostname
- e1.jobID as jobID
- e1.jobName as jobName
- e1.metricsName as metricsName
- e1.metricsKey as metricsKey
- e1.metricsValue as metricsValue
- e1.tmlID as tmlID
- e1.daemon as daemon
- e2.metricsKey as finalKey

# 机器学习-在线训练

应用管理 /

退出编辑

保存画布

数据源

数据预处理

特征工程

算法

验证

输出

停止

重启

模型部署

资源

配置

告警

版本

MXNet-FM

模型验证

模型输出到HDFS

MXNet-FM

标签字段名

label

ml.feature.fields

单次迭代训练数据数目

32

优化算法

SGD

学习率

0.1

动量

0

梯度更新范围

0

分解因子向量维度

2

线性项L1系数

1

线性项L2系数

1

因子项L1系数





0

因子项L2系数

0



# 机器学习-在线推理



模型服务

应用管理 / 模型服务

部署时间: 2019-07-13 00:00:00 ~ 2019-07-14 23:59:59

输入关键字进行搜索

模型服务组	模型生成时间	当前状态	责任人	部署时间	运行时长	最后操作	操作
serving001	2019-04-30 11:00:13	运行中	admin	04-30 14:57	2 个月	robbinli	停止 测试 评估报告 日志 删除

< 1 >



THANK YOU!

